Three Essays in Applied Microeconomics

Konrad Burchard Burchardi

Dissertation submitted for the degree of

Doctor of Philosophy in Economics at

The London School of Economics and Political Science

June 2011

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others. The first chapter draws on work that was carried out jointly with equal share by Tarek A. Hassan and me. The third chapter draws on work that was carried out jointly with equal share by Stefan P. Penczynski and me.

Konrad Burchardi Burchardi

Abstract

This thesis comprises three independent chapters, spanning the range of my interests. The first chapter provides estimates of the causal effects of social ties on aggregate, firm level and individual level economic outcomes. The second chapter is a first step at understanding the joint determination of language change and economic structural change which seems to have occurred over the past centuries and probably continues today. A joint theme in these papers is the attempt to contribute to our understanding of the extent to which cultural and social factors can impact market outcomes. The second paper is as well interested in a specific channel through which economic outcomes can feed back on a cultural aspect of societies. The third paper seeks to contribute to our understanding of a more traditional economic question, namely individual behaviour in strategic situations. In particular it uses a novel experimental design to investigate individual behaviour in unprecedented strategic situations and estimate the parameters of a structural non-equilibrium model of behaviour. I like to belief that the chapters of this thesis, especially the second and third chapter, tie up closely the theoretical and empirical work and make a humble contribution to our understanding of economic behaviour and market functioning.

Für meine Eltern, Ruth und Klaus Burchardi

Acknowledgements

Reflecting on whom I want to thank, I come to realise that completing the PhD leaves me indebted to many people. It is a comfortable feeling that so many people have helped and supported me during the past years.

Firstly, I am greatly indebted to my supervisors, Tim Besley and Maitreesh Ghatak. They have supported me throughout the past years, provided guidance and encouragement when necessary and showed a level of trust and belief in me, that was – while comforting and encouraging – at times surprising. I have learned a great deal from them and I have been impressed by them. I much enjoyed working for and with them, and they have shaped my understanding of our discipline and beyond. I will keep a dear memory of the many discussions and I am grateful for their support.

I am as well indebted to Stefan P. Penczynski and Tarek A. Hassan with whom I have worked on the third and first chapter of this thesis, respectively. Stefan has been a close friend during the years in London and at LSE. We have also coincided in visiting Universitat Pompeu Fabra, Barcelona, between April and August 2007. Most of the initial work for the third chapter was done during this time, and Stefan and I shared the view, that this can be equally well done – if not better – in cafés, in parks, or on white sand. In fact, we had the key idea for the experimental design when sitting on the Mediterranean. I have met Tarek thanks to the U.S. Department of Homeland Security (to whom I do not feel indebted) and I have benefited substantially from working with and by learning from him. He invited me to Cambridge and arranged for me to present at Harvard, which – even if my academic career fails, or especially then – I will tell my potential grand-children more often than they will want to hear. Tarek has been an incredibly supportive friend and co-author during my job applications.

My time at LSE has been an extraordinarily inspiring and enjoyable experience. During my first year here, while completing the M.Sc. in Economics, I have enjoyed the enthusiasm for understanding (and changing) the world of many impressive fellow students. I continue to share friendships with Marco Alfano, Fernando Aragon, Richard Bradley, Roy Havemann, Matt Lang, Declan Magee, and Ona Vilanou i Figueras, amongst others. When joining the Ph.D. programme in the following year, I was privileged to become a member of EOPP/STICERD. Probably the single most important factor for my experience at LSE were all those who made STICERD such a stimulating research environment. I have learnt from and enjoyed the numerous and endless discussions about economics and just about everything with Madhav Aney, Fernando Aragon, Erlend Berg, Patrick Blanchenay, Kara Contreary, Dave Donaldson, Oliver Denk, Nathan Foley-Fisher, Jason Garred, Pedro Gomes, Sónia Gonçalves, Selim Gulesci, Giovanni Ko, Rocco Macchiavello, Elena Manzoni, Inés

Moreno de Barreda, Ameet Morjaria, Hannes Müller, Sanchari Roy, Stefanie Sieber, Miriam Sinn, and Dimitri Szerman. Many of them became friends who were always up for a chat when I was bored or tired – even if they were just chasing the idea of the term. The same is true for my dear friend Mariano Cena, who has an admirable ability to put things into perspective.

Many individuals have provided feedback on the chapters of this thesis. The first chapter has benefited from comments by Philippe Aghion, Oriana Bandiera, Robin Burgess, Davide Cantoni, Wendy Carlin, Greg Fischer, Claudia Goldin, Jennifer Hunt, Erik Hurst, Nathan Nunn, Rohini Pande, Torsten Persson, Michael Peters, Steve Pischke, Steve Redding, Daniel Sturm, and seminar participants at Harvard, the University of Chicago, UCLA, the London School of Economics, the EDP Meeting London, London Business School, Warwick, MPI Bonn, Carlos III, IIES, Northwestern Kellogg, UCLA Anderson, Universitat Pompeu Fabra, Bocconi, CEMFI and the EBRD. I am grateful to Jason Garred for carefully proof-reading and commenting on a full draft of the paper and Daniel Sturm's feedback during the job application process. The second chapter of the thesis has benefited from comments from seminar participants at LSE and I am especially indepted to Tim Besley for his ongoing encouragement to proceed with this project. The third chapter has benefited from comments by Madhav Aney, Ayala Arad, Dan Benjamin, Colin Camerer, Mariano Cena, Syngjoo Choi, Vincent Crawford, Ziad Daoud, Erik Eyster, Gilat Levy, Michele Piccione, Mireia Raluy i Callado, Ariel Rubinstein, Reinhard Selten, and seminar audiences at Cornell, ESA International Meeting 2009, ESA North America Meeting 2009, GfeW Meeting 2008, LSE, Nordic Behavioural and Experimental Conference 2008, Universitat Pompeu Fabra, and University of East Anglia. Kara Contreary, Jan Zapal, and Piotr Zurawski provided excellent research assistance. Part of the work was accomplished at the Universitat Pompeu Fabra in Barcelona, Spain, and I benefited from its hospitality, especially by Rosemarie Nagel and Nagore Iriberri. I would like to thank Georg Weizsäcker for his encouragement to start this project and many helpful discussions.

I gratefully acknowledge financial support during my PhD from the Adeline and Karl Goeltz Scholarship Programme, the ESRC, LSE, the Royal Economic Association and the C & J Modi / Narayanan Fellowship. The work on the third chapter has been financially supported by the Michio Morishima fund (STICERD), the London School of Economics and the German Association of Experimental Economists (GfeW, Heinz-Sauermann Förderpreis 2007).

Most importantly, I am grateful to my parents and my elder brothers, Christian and Klaus-Henning, for their love, trust, care, support and many things I lack words for and to Mireia for her love, patience, and understanding. I should have spend more time with all of them.

Contents

List of Figures

List of Tables

Preface

This thesis comprises three independent chapters, spanning the range of my interests. The first chapter provides estimates of the causal effects of social ties on aggregate economic outcomes. Little causally interpretable evidence on this topic is available, the reason being that social ties are generally determined jointly with or by economic outcomes. The chapter shows that at the time of the German Reunification there was substantial variation across West German regions in the share of individuals with ties to East Germany. We argue that there is an exogenous reason for this pattern and we make this explicit in our empirical strategy. We find that regions in which a higher share of the population had ties to East Germany experienced a significantly higher growth in income per capita in the 6 years subsequent to the fall of the Berlin Wall. Further we find that firms from these regions are more likely to hold subsidiaries or branches in East Germany even today and that the share of entrepreneurs in these regions increased differentially. Lastly, we find that individuals with ties to East Germany experience a significant rise in their income. However, the sum of the individual level effects is smaller than the regional level effects, suggesting some form of regional spill-over effects to individuals who themselves do not have ties to East Germany, but live in a region where a high fraction of the population has ties. We believe the most plausible interpretation of the sum of our results is that West German households which had social ties to East Germany in 1989 had a comparative advantage in seizing the new economic opportunities in the East. This comparative advantage resulted in a persistent rise in their individual household incomes but also appears to have generated a social surplus at the regional level in the form of higher growth in income per capita and increased returns to entrepreneurial activity. Part of this social surplus may be explained if firms owned by a household with social ties to the East (or firms who had access to a local labor force with such ties) had a comparative advantage in investing in East Germany.

The second chapter of this thesis is an attempt at contributing to our understanding of the co-evoluation of linguistic heterogeneity and economic structural change. Today, linguistic heterogeneity at the national level, which is a widespread phenomenon, correlates negatively with the level of GDP and the fraction of the population which is working in manufacturing and services. Historically, much of the now highly industrialized world was home to much greater linguistic variety, too, but the usage of many local lanaguages declined at around the time and subsequent to the industrial revolution. These facts give rise to the sense that the evolution of language heterogeneity, economic growth and structural change in the economy might be connected. The chapter aims to provide a framework which helps to understand the two-way relation of language knowledge and the sectorial composition of the economy and economic growth. In particular, it formalizes how the stock of language knowledge restricts the set of feasible joint economic activities, and how these in turn shape the incentives for language change. The model highlights two key structural parameters which can be interpreted as answering the questions: 'How important is a common language for economic activity?' and 'How important are economic incentives for language change?' The model's structural equations are used to estimate these parameters. I calibrate the model's remaining parameters with 1926 data from French départements and simulate the general equilibrium model to obtain predictions on language heterogeneity and sectoral composition across départments in 1946. I show that the model of language change and sectorial change explains at least 7% of the variation in the sectorial composition.

The third chapter investigates individual behaviour in unprecedented strategic situations. Previous experimental evidence shows that equilibrium predictions often fail to explain behaviour in one shot games, which are meant to represent unprecedented strategic situations. A leading explanation for the failure of equilibrium concepts to explain and predict behaviour in one shot games is that players may not believe that other players choose an equilibrium strategy. This explanation appears particularly pertinent in unprecedented strategic situations, where learning cannot cause a convergence of beliefs and strategies to equilibrium. The level-k model of reasoning postulates the following alternative belief structure: There exist so-called level-0 players, who do not play strategically. The model defines level-1 players to best respond to what they believe level-0 players do. Level-2 players form a belief about the fractions and strategies of lower level players and best respond to this. This process continues for higher level players. Hence the model assumes a hierarchy of types who best respond to non-equilibrium beliefs, referred to as *levels*. They distinguish themselves by the number of iterated best responses to the distribution of level-0 actions. While there are various experimental studies which aim to estimate the distribution of strategic players, little in known about the existence of non-strategic players and their behaviour. The third chapter aims to contribute to our understanding of these questions. It presents a novel experimental design which allows to obtain incentivised written accounts of individuals' reasoning. Contrary to commonly held beliefs, we show that around one third of the participants play non-strategically. And while the non-strategic actions are not uniformly distributed, higher level players, somewhat surprisingly, correctly anticipate the non-uniform distribution of non-strategic actions.

1 The Economic Impact of Social Ties: Evidence from German Reunification[1]

1.1 Introduction

There are important theoretical reasons to believe that social ties between individuals impact economic development. Social ties might facilitate communication and thereby reduce informational frictions and asymmetries (Granovetter (1973), Varian (1990), Stiglitz (1990)). Furthermore, the threat of severing social ties may serve as a form of 'social' collateral and thereby sustain a large range of credit, insurance and trade contracts that would not otherwise be feasible (Besley and Coate (1995)). Although both of these channels may fundamentally affect the ability of individuals to engage in economic transactions, there exists virtually no evidence to date on the relevance of social ties for aggregate economic outcomes, such as economic growth and the scale and success of entrepreneurial activity. In this paper we estimate the relevance of social ties between individuals for regional economic development and trace the economic impact of social ties to firm investment and household income.

The main obstacle to estimation of the causal effect of social ties on economic outcomes is that social ties are endogenous to economic activity. At the microeconomic level, individuals may form social ties in anticipation of future economic benefits, or as a result of economic interaction. At the aggregate level, the regional distribution of social ties is a result of decisions of individuals about where to live, and these decisions are endogenous to economic incentives. Identifying a causal link between social ties and economic outcomes thus requires (i) the identification of 'real friends', i.e. social ties that formed without regard to future economic benefits, and (ii) some exogenous variation in the regional distribution of these exogenously formed social ties. In this paper we use the fall of the Berlin Wall as a natural experiment, which enables us to overcome both of these difficulties by exploiting two peculiar features of Germany's post-war history. First, the partition of Germany was generally believed to be permanent, so that individuals maintaining social ties across the inner-German border must have done so for purely non-economic reasons. Second, the pattern of wartime destruction in West Germany made it temporarily much more difficult to settle in some parts of West Germany than in others, precisely during the period during which millions of refugees and expellees from the East arrived in the West.

We find that West German regions which (for exogenous reasons) received a larger inflow of expellees from East Germany before the construction of the Berlin Wall have significantly stronger social ties to East Germany in 1989 and exhibit sub-

[1] The work in this chapter was carried out jointly with equal share by Tarek A. Hassan and me.

stantially higher growth in income per capita after the fall of the Berlin Wall. This economic expansion is associated with a rise of the returns to entrepreneurial activity and with an increase in the share of the population who are entrepreneurs. Moreover, firms which are headquartered in West German regions which have stronger social ties to East Germany in 1989 are more likely to operate a subsidiary or a branch in eastern Germany today. In addition, West German households who have a relative in East Germany in 1989 experience a persistent rise in personal income after the fall of the Berlin Wall. The household level effects explain only around one sixth of the aggregate level effects, suggesting strong spill-over effects. We interpret these findings as evidence of a causal link between social ties and regional economic development in West Germany.

The first key advantage of the natural experiment surrounding German reunification is the fact that the fall of the Berlin Wall was largely unexpected. After the physical separation of the two German states in 1961, private economic exchange between the two Germanys was impossible.[2] Social ties that West Germans maintained with East Germans during this period were then kept up for purely non-economic reasons, as individuals on both sides of the border did not expect an economic re-integration. On November 9th 1989, these social ties suddenly took on economic value: after the fall of the Berlin Wall, trade between the two Germanys became feasible, following more than three decades of isolation. The result was a boom in economic exchange between West and East.[3] In this situation, East Germans had valuable local information about demand conditions and about the quality of the assets that were offered to investors. However, they were largely unable to borrow (indeed, until the mid-1990s many did not know whether they owned their own homes) and lacked experience of the rules and norms of behavior in a capitalist economy. West Germans, on the other hand, had these capacities but lacked the requisite local knowledge. To the extent to which social ties facilitate economic exchange, social ties between East and West Germans thus suddenly took on economic value on the day of the fall of the Berlin Wall.

In our analysis we work with aggregate data on the share of individuals with social ties to East Germany.[4] This dataset shows that there was substantial variation in the share of individuals with social ties to the East across West German regions. The second key advantage of the natural experiment surrounding German reunification is that much of this variation has its roots in the idiosyncrasies of Ger-

[2]In fact it had been highly restricted as far back as the late 1940s.

[3]Moreover, almost the entire East German capital stock was sold to private investors between 1990 and 1994. This boom was fueled by large transfers from West to East. These included both direct and indirect government transfers. For example, the East German Mark was converted to the Deutsche Mark at several times its market value. See (Sinn and Sinn, 1992, p. 51) and Lange and Pugh (1998), respectively.

[4]This measure of social ties thus captures bilateral ties. We are therefore not able to test network-theoretical models of social ties.

many's post-war history. Indeed, we show that West German regions which received a large inflow of *refugees* and *expellees* from East Germany between 1945 and 1961 tend to have significantly stronger social ties to the East in 1989.[5] However, the assignment of refugees and expellees from the East to West German regions might not be random, as individuals may have moved to those regions in which they saw the best prospects for themselves and their families. In particular, an overwhelming concern for those arriving from the East after 1945 was the acute housing crisis in West Germany, which persisted until the early 1960s. This crisis resulted from the fact that during World War II, 32% of the West German housing stock was destroyed. The expellees and refugees arriving from the East were thus channeled predominantly into areas in which there was relatively more intact housing, i.e. to the areas that were least destroyed during the war.

Variation in wartime destruction which temporarily made it more difficult to settle in some parts of West Germany than in others thus provides the exogenous source of variation in the regional distribution of social ties which we need in order to identify a causal effect of social ties on regional economic outcomes. In particular, we use the degree of wartime destruction in 1945 as an instrument for the share of expellees settling in a given West German region. The main identifying assumption for our region-level results is thus that the degree of wartime destruction in 1945 (or any omitted factors driving it) affected growth in income per capita after 1989 only through the settlement of refugees and expellees post World War II; and that these groups indeed affect growth post-1989 exclusively due to their social ties to East Germany.

We devote a great deal of care to corroborating this identifying assumption in various ways. For example, we show that wartime destruction is uncorrelated with pre-war population growth, that it affects post-war population growth only until the 1960s, and that it has no effect on the growth of income per capita in West German regions in the years before 1989. Moreover, all of our specifications are robust to controlling for the growth in income per capita in the years prior to 1989. Finally, we show that our results are particular to expellees arriving from East Germany rather than to other expellees who arrived directly to West Germany, i.e. without settling in East Germany for some time.

Based on this identifying assumption, we show a strong causal relationship between the intensity of social ties to East Germany and post-1989 growth in income per capita among West German regions: a one standard deviation rise in the share

[5]'*Refugees*' are individuals who in 1939 had their primary residence in the territory that became the Soviet sector post-World War II and who then fled to West Germany between 1945 and 1961. '*Expellees*' are individuals who in 1939 had their primary residence in areas of the German Reich which became part of Poland, Russia or another country from which ethnic Germans were expelled post-World War II. We focus on the group of expellees who held a residence in the Soviet sector after the war, but migrated to West Germany between 1945 and 1961 (as did the vast majority of expellees who were originally allocated to the Soviet sector).

of expellees from East Germany settling in a given West German region in 1961 is associated with a 4.6% rise in income per capita over the six years between 1989 and 1995. This is a sizable effect, amounting to around 0.7 percentage points higher growth per year.[6] While the regional growth effect diminishes after 1995, there is no evidence of a subsequent reversal, so that the pattern of social ties to the East which existed in 1989 may have permanently altered the distribution of income across West German regions.

In an effort to shed light on the mechanism linking social ties to regional economic growth we estimate separate effects for the incomes of entrepreneurs and non-entrepreneurs. While both entrepreneurs and non-entrepreneurs who live in regions with strong social ties to the East experience a significant rise in their incomes, the incomes of entrepreneurs increase at more than twice the rate of those of non-entrepreneurs. Consistent with this observation, the share of the population engaged in entrepreneurial activity rises in regions with strong social ties.

To trace this effect of social ties on entrepreneurial activity to the behavior of firms, we use data on the cross-section of West German firms in 2007. We show that West German firms which are headquartered in a region which had strong social ties to the East in 1989 are more likely to operate a subsidiary or a branch in East Germany in 2007. In particular, a one standard deviation rise in the share of expellees from East Germany settling in a region before 1961 is associated with a 3.4% increase in the likelihood that a given firm within that region operates a subsidiary or branch in East Germany in 2007. Interestingly, this effect seems to be concentrated in the services sector, which is consistent with the view that social ties are particularly important in industries which rely heavily on local information. While social ties to East Germany predict a higher probability of investing in East Germany, they do not predict a higher probability of investing anywhere else in the world, except for a small rise in the probability of investing in Poland. This latter finding is notable as many of those arriving in West Germany from East Germany before 1961 were originally expelled from present-day Poland in 1945, then lived in East Germany for up to 16 years and arrived in the West before the construction of the Berlin Wall in 1961.

We then estimate the effect of social ties at the household level. We show that West German households who report having relatives in East Germany in 1989 experience a persistent rise in their income after the fall of the Berlin Wall: the income of households with at least one relative in the East rises on average by 4.9% over the six years following the fall of the Berlin Wall. This rise in income again occurs immediately after 1989 and is robust to controlling for possible omitted

[6]This result is robust to several different variations in the estimation strategy and cannot be explained by likely alternatives, such as migration from East to West in 1989 or by a decline in West German heavy industry in the 1990s.

variables, such as the age, level of education, or capital income of the household head. Especially households headed by individuals below the age of 40 (lowest quartile of the age distribution) and above the age of 52 (top two quartiles) seem to profit from their social ties to East Germany. We also show that households profit from their ties to East German relatives regardless of their level of capital income in 1989, which is consistent with the view that households used their ties to the East primarily as conduits for information, rather than to facilitate borrowing.

We interpret our results as evidence that social ties to East Germany indeed took on significant economic value after the fall of the Berlin Wall: West Germans who had ties to the East were better able to take advantage of the new economic opportunities in the East, possibly because they were better informed about investment opportunities than their peers. The rise in regional income seems to be driven primarily by an increase in entrepreneurial activity and by an increase in the number of entrepreneurs. Moreover, firms which had access to a workforce with strong social ties seem to have had a comparative advantage in investing in the East. Comparing the quantitative implications of our household-level and our region-level results suggests that there were significant spill-overs, through which households living in regions with strong ties to the East experience rising incomes, even if they themselves do not have direct personal ties to the East.

While we believe that this paper convincingly demonstrates the relevance of social ties for regional economic development, there are two important caveats to the interpretation of our results. First, it is unclear how our results generalize beyond the context of a large economic transition, such as the economic re-integration of Germany. Social ties may be particularly useful in an environment in which markets are established rapidly and informational asymmetries are large. Second, we cannot be sure whether social ties led to an increase of economic activity at the country level or whether they merely resulted in a re-distribution of rents from regions with weaker social ties to regions with stronger social ties. However, the fact that we find positive spill-overs rather than crowding out at the regional level suggests that Germany as a whole was better off for its access to individuals with social ties.[7] Finally, some of the patterns that we document may be explained if there remain unobserved cultural differences in 1989 between 'native' West Germans and the population of expellees and refugees who settled in West Germany post-World War II. We do not emphasize this interpretation, as Germans of all parts were fairly homogenous before the separation, sharing a common language and culture. The integration of the new arrivals into West German society is commonly regarded as one of the preeminent achievements of post-war Germany, such that people would

[7] In the main part of the paper we only estimate the effects for West German districts - mainly due to a lack of East German data before 1989. If some of the surplus generated by social ties accrued to East Germans, our estimates likely underestimate the total growth effect of social ties.

typically not know (or care) whether their neighbors were descendants of expellees or refugees after the war.

To our knowledge, this paper is the first to identify the effect of social ties on aggregate economic outcomes and the first to trace this effect from the households and firms to regional economic development. Our results relate to a large literature which links social networks and social ties to a broad set of microeconomic outcomes, ranging from employment (Munshi (2003), Laschever (2007), and Beaman (2008)) and informal insurance (e.g. Weerdt and Dercon (2006)) to performance in the financial industry (Cohen, Frazzini, and Malloy (2008), Hochberg, Ljungqvist, and Lu (2007), and Kuhnen (2009)) and agricultural yields (Conley and Udry (2009)). Since social ties have been documented to influence such a wide range of microeconomic outcomes, an obvious question to ask is whether they also influence aggregate economic variables. While there are number of models that predict such aggregate effects (e.g. Rauch (1999), Rauch and Casella (2003), Kranton and Minehart (2001), and Ambrus, Mobius, and Szeidl (2010)), we believe our paper to be the first to provide causally interpretable evidence on this matter.

Most closely related to our empirical work is the paper by Fuchs-Schündeln and Schündeln (2005) who use the reunification of Germany to identify the role of risk aversion in occupational choice. A number of other authors have used the partition of Germany as natural experiment. Redding and Sturm (2008) estimate the effect of market access on city growth, Alesina and Fuchs-Schündeln (2007) estimate the effect of the East German regime on voter preferences, and Bursztyn and Cantoni (2009) estimate the effect of exposure to West German media on consumer preferences.

In using the effect of wartime destruction on the settlement of expellees in West Germany, we also relate to a number of papers which quantify the effect of wartime destruction on long-run economic development. At the aggregate level, Brakman, Garretsen, and Schramm (2004) find that the pattern of wartime destruction had no long-run impact on city growth in West Germany. Davis and Weinstein (2002) and Miguel and Roland (2011) show similar results for Japan and Vietnam, respectively. At the individual level, Akbulut-Yuksel (2009) documents a detrimental effect of wartime destruction on the education and health of school-age children who grew up in the most affected areas.

The remainder of the paper is organized as follows: Section 1.2 reviews the historical background of wartime destruction, the partition of Germany, and the settlement of expellees and refugees in West Germany. Section 1.3 discusses the data and its construction. Section 1.4 establishes the basic relationship between wartime destruction and social ties to the East, and uses this relationship to identify a causal effect of social ties on growth in income per capita post-1989. Section 1.5 provides evidence which suggests that entrepreneurial income and activity being the main driver of aggregate income growth. In particular, section 1.5.2 documents the influence of social ties on the ownership structure of German firms today. Section 1.6 looks at the relationship between social ties and income growth at the household level. Section 1.7 discusses possible mechanisms by which social ties might affect economic activity. Section 1.8 discusses our results, while the appendix contains additional robustness checks and details on the construction of our dataset.

1.2 Historical Background

1.2.1 Destruction of Housing Stock during World War II

German cities and towns were heavily destroyed after World War II. This was mainly the result of Allied air raids, which began in 1940 and intensified until the final days of the war in 1945. These left around 500,000 dead and resulted in the destruction of a third of the West German housing stock, making it the most devastating episode of air warfare in history.[8]

In the early days of the war the Royal Air Force attempted to slow down the advance of the German army into the Soviet Union by destroying transport infrastructure. This strategy was an abject failure and was quickly abandoned, as the available technology at the time did not permit targeted raids. At best, the pilots flying the nighttime raids were able to make out that they were above a city (and they were often even unsure which city lay below). This led to the adoption of the doctrines of 'moral bombing' (1941) and of 'fire and carpet bombing', which were aimed at destroying the Germans' morale and ability to resist by destroying cities and towns (Kurowski (1977)). By the end of the war, 50% of the 900,000 metric tons of bombs deployed had hit settlements, while 17% had hit industry or infrastructure.

The cities destroyed most during the early years of the war were those that were close to the British shore and easy to spot from the air, e.g. Hamburg and Cologne. After 1944, utilizing recent technological advances, the allies were able to implement fire storms, which were easiest to create in cities with highly flammable, historical centers, such as Darmstadt, Dresden, or Wuerzburg. Fire storms could typically not be implemented in cities which had already been hit by a large number of explosive bombs, as the rubble from earlier raids would have prevented the fire from spreading. This is why the cities that were attacked relatively late in the war (often strategically the least important) were among the most heavily destroyed.[9]

Figure 6 shows the varying intensity of destruction in West German regions. Note that none of our empirical results rely on this pattern being random or driven by certain factors. Instead, our identification strategy relies on the assumption that

[8]The information presented in this section is from U.S. Govt. Print Office (1945), Kurowski (1977), and Friedrich (2002).

[9]During a fire storm, a large section of a city catches fire, creating winds of up to 75 meters per second, depriving those exposed of oxygen and often sucking them into the fire.

the pattern of wartime destruction or any omitted factors driving it have no direct effect on growth in West German regions 45 years later, post 1989.

1.2.2 The Partition and Reunification of Germany

In 1944, as World War II entered its final phase, the UK, the US and the Soviet Union agreed on a protocol for the partition of pre-war Germany: The areas to the east of the rivers Oder and Neisse were to be annexed by Poland and by the Soviet Union, and the remaining territory was to be divided into three sectors of roughly equal population. The UK would occupy the Northwest, the US the South, and the Soviet Union the East. The capital, Berlin, would be jointly occupied. At the end of the war, the three armies took control of their sectors, and the US and Britain carved a small French sector out of their territory.[10] In 1949, with the onset of the Cold War, the three Western sectors formed the Federal Republic of Germany (West Germany), and the Soviet sector became the German Democratic Republic (East Germany). Economic exchange between the two parts of Germany became increasingly difficult as the East German government immediately introduced central planning. Only three years later, in 1952, the border was completely sealed, cutting any remaining legal or illegal trade links between East and West.[11]

Until the construction of the Berlin Wall in August 1961 there remained the possibility of personal transit from East to West Berlin, which was the last remaining outlet for refugees fleeing from East to West Germany. After 1961, migration between East and West virtually ceased. In the following years the partition of Germany was formally recognized in various international treaties, and was, until the summer of 1989, generally believed to be permanent.[12]

In September of 1989 it became apparent that a critical mass of East Germans had become alienated from the socialist state, its declining economic performance, and the restrictions it placed on personal freedom. Increasingly large public demonstrations led to the opening of the Berlin Wall on November 9th, 1989. The first free elections in East Germany were held in March of 1990, followed by the rapid political, monetary, and economic union between East and West Germany by the end of the same year.

[10]This entailed a significant withdrawal of the British and US forces, who had captured much more territory than expected (Sharp (1975)).

[11]The only remaining trade between the two countries was the 'Interzonenhandel' which was arranged between the two governments. In this system the East German government would trade goods and services by the barter system. In 1960 its total volume came to the equivalent of $178 m. See Holbik and Myers (1964) for a detailed description of the Interzonenhandel.

[12]The most important of these treaties was the 'Grundlagenvertrag' of December 1972 between East and West Germany in which both countries recognized 'two German states in one German nation.' Following this treaty East and West Germany were accepted as full members of the United Nations.

1.2.3 Refugees and Expellees in Western Germany, 1945-1961

In 1945 the Polish and Soviet authorities expelled all German nationals from the annexed territory, so that areas which used to be German before 1939 could be inhabited by Polish (and Russian) nationals after 1945. While many Germans in these regions had fled the advancing Soviet Army, those that remained were marched or transported out of the annexed territories towards the four sectors. We refer to this group of people as 'expellees'. Germans that either originally lived in or moved to the countries occupied by the German army during war were also expelled in many cases, particularly from Czechoslovakia, Hungary, Romania, and Yugoslavia. Expellees were registered and then assigned one of the four sectors in which to settle, according to quotas fixed in the Potsdam accord.[13] The authorities in turn allocated the expellees to the (later to become federal) states within their jurisdictions and assigned them quarters wherever they could find intact housing stock.

The first wave of 5.96 million expellees arrived in the three Western sectors by 1946. We refer to this group as 'direct expellees'. As it became increasingly apparent that the division of Germany would become permanent, most of the 3.04 million that had originally been allocated to the Soviet sector left for the West. These 'expellees via the Soviet sector' are critical to our empirical analysis as they had the opportunity to form social ties to East Germans before migrating on to West Germany. By 1960, the total number of expellees in West Germany had risen to 9.697 million, of which roughly one third were expellees via the Soviet sector.[14]

In parallel, an increasing number of native residents of the Soviet sector who were dissatisfied with the political and economic prospects of the fledging East Germany fled to the West. This flow of 'refugees' peaked in the years before the construction of the Berlin Wall, with on average around 300,000 individuals illegally crossing the Border in each year between 1957 and 1961 (Hunt (2006)). By 1961 the total number of East German refugees settling in West Germany was 3.5 million.

While the authorities in the western sectors, and later the West German authorities, had an explicit policy supporting expellees, supplying them with housing and various subsidies, there was very little support for refugees. In fact, as late as 1950 the authorities actively tried to discourage refugees from entering West Germany on the grounds that they would exasperate an already catastrophic situation in the housing market and for fear of the political consequences of a de-populating East Germany. In practice, however, the authorities never attempted to deport refugees back to the East, and so refugees often made their own way in West Germany,

[13]The official plan adopted by the allies in November 1945 was to expel 6.65 million Germans. 2.75 million, were to be allocated to the Soviet sector and 2.25 million, 1.5 million, and 0.15 million to the American, British, and French sectors, respectively (Bethlehem (1982), p.29).

[14]We are unable to determine exactly how many expellees remained in East Germany, as the communist government declared after 1950 that the expellees had been fully integrated into East German society and banned the concept from subsequent government statistics (Franzen (2001)).

without registering with the authorities.[15] The severe housing crises that resulted from the inflow of millions of migrants into the heavily destroyed Western sectors remained the principal determinant in the allocation of expellees and refugees to West German cities and towns until the late 1950s.[16]

1.3 The Data

We use data at the household, firm, district (Landkreis), and regional (Raumordnungseinheit) level. Districts are the equivalent of US counties. Regions are the union of several districts, and each district belongs to one such unit. Regions do not have a political function but exist exclusively for statistical purposes; in this sense they are analogous to metropolitan statistical areas in the US, but also encompass rural areas. All of our aggregate data is available at the district level, except for income per capita before 1995, which is available only at the regional level. Our primary units of analysis are thus the 74 West German regions. When we use aggregate controls in our firm and household level analysis we always use data at the lowest level of aggregation available.

1.3.1 Region Level Data

Our primary proxy of social ties to the East at the region level is the share of expellees via the Soviet sector in 1961. The 1961 census reports the number of inhabitants and the number of expellees in each West German district. The census presents the data separately for expellees who arrived directly in West Germany in or after the war and those who arrived in West Germany after having registered a residence in the Soviet sector. From this data we created the variables *Share Expellees (Direct) '61* and *Share Expellees (Sov. Sector) '61*.[17]

For our instrumental variables strategy we coded two measures of wartime destruction: the share of dwellings that were destroyed in 1946, labeled *Share Housing Destroyed '46*, and the amount of rubble in cubic meters per inhabitant, labeled

[15] See Bethlehem (1982, chapter 3).

[16] In the early years the availability of housing was the only determinant of where the expellees were sent (Bethlehem, 1982, p. 29, pp.49). After 1949 economic considerations started playing a more important role in the allocation process and the West German government also initiated a number of programs encouraging migration to areas in which there was a relatively higher demand for labor. However, these programs remained relatively limited, with less than one in ten expellees participating in them.

[17] This data was collected at the 1961 district level. Some West German district boundaries have changed between 1961 and 1989. Our analysis uses the 1989 district boundaries. To calculate the district level share of expellees on the basis of the 1989 district boundaries, we proceeded as follows: We first tracked (using ArcGIS) which share of the area of each 1961 district became part of which 1989 district. For example, we calculated that 71% of what used to be district 'Mainburg' in 1961 became part of a district called 'Kelheim' in 1989, and the remaining 29% became part of the 1989 district 'Freising'. Assuming that the expellees were distributed homogenously across space within districts in 1961, we calculate the number of expellees living in each 1989 district.

Rubble '46 (m^3 p.c.). Both measures are from the 1946 edition of the annual statistical publication of the German Association of Cities. This data is reported at the city level for the 199 largest West German cities and towns. We also coded the number of inhabitants of these towns in 1939 and 1946 from this volume. We aggregated the data on wartime destruction by calculating the mean destruction across cities in a district or region, weighted by the cities' population in 1939. Additional details on the data sources used are given in Appendix A.3.

Our data on income per capita is from the German 'Mikrozensus', an annual, obligatory random survey of one percent of the population. We aggregated the individual income data to the region level for every second year between 1985 and 2001. Income per capita at time t is labeled *Income t (p.c.)*. As the 'Mikrozensus' does not identify districts prior to 1995, an aggregation to districts was not possible. We also used the 'Mikrozensus' to construct the average income of entrepreneurs (*Income (p.c.) Entrepreneurs t*), the average income of all others (*Income (p.c.) Non-Entrepreneurs t*), as well as the share of entrepreneurs amongst the respondents (*Share Entrepreneurs t*) for each region. From the same source we obtained data on the share of the population working in the manufacturing sector in 1989 (*Share Working in Manufacturing '89*) and the share of the population that migrated to the region from the East in the years 1991, 1993 or 1995 (*Migration from East '91-'95*). As an additional control we calculated the distance of the center of each district or region to the former inner German border from GIS data (*Distance to East (100 km)*).

1.3.2 Firm Level Data

Our firm level data is from the 2007 edition of the ORBIS dataset. This source contains data on some 750,000 firms in West Germany and their subsidiaries and branches both in Germany and abroad, including information on the postal code of the firms' headquarters and of their associated subsidiaries and branches. We used this information to match each firm to the West German district in which its headquarters are located. We selected all firms which have at least one subsidiary or branch in a West German district other than the district of the headquarters ($N = 19,420$). As a simple measure of firms' investment activity in different parts of the world, we created a dummy variable for whether the firm has a subsidiary or branch in location x (*S. & B. in x (Dummy)*), made up of separate variables for 'East Germany', 'Poland', the 'Old EU Countries', the 'New EU Countries (excluding Poland)' and for 'Non-EU Countries' ('Old EU Countries' refers to the 14 EU member countries other than Germany prior to enlargement in 2004). For the same set of firms we computed the share of firm's subsidiaries and branches in location x as a fraction of its total number of subsidiaries and branches in location x and West

Germany (*Share of Total S. & B. in x*). As proxy for the size of the firm we use the number of subsidiaries and branches it operates in West Germany (*S. & B. in West Germany*). Finally, we used the NACE code given in the ORBIS dataset to define four sectoral fixed effects (agriculture, manufacturing and construction, trade and service, and government).

1.3.3 Household Level Data

Our household level data is from the German Socio-Economic Panel (SOEP), which is an annual panel of German households. From the panel we selected households which participated in the 1985, 1989, and 1995 waves, and used information on household income in the years 1985-2001 (*Income (SOEP)*), the amount of capital income in 1989, and the age and years of education (including professional education) of the household head. We also created dummies for the primary occupation and gender of the household head.[18]

Importantly, the SOEP questionnaire asked in 1991 whether the respondent had any relatives in the other part of Germany. From this information we constructed our measure of social ties at the household level (*Ties to Relatives '91*), which is a dummy variable equal to 1 if at least one individual in the household had a relative in the other part of Germany. We also aggregated this variable to the region level by calculating the share of households with ties to East Germany in each West German region (*Share Ties to Relatives '91*), which we use as a secondary measure of social ties. Lastly, we created a dummy variable indicating households whose household head was an entrepreneur in 1989 (*Entrepreneur '89*).

1.3.4 Descriptive Statistics

Table 1 provides descriptive statistics. Panel A of Table 1 presents the data on West German regions; here, the second and third columns divide the sample into regions with a higher and lower share of housing destroyed in 1946 than the median region. The first row of column 1 gives the mean and the standard deviation of the share of expellees via the Soviet sector in 1961, while the second column gives the mean and standard deviation for regions with below median levels of wartime destruction. Expellees via the Soviet sector made up 4.8% of the 1961 population in the average region. Similarly, expellees that came directly to West Germany made up 11.9% of the average region's population in 1961 (row 2), and 22.3% of the population report having relatives in East Germany in 1991 (row 3). In all three cases, these shares are higher in regions that suffered lower levels of wartime destruction. The variation in wartime destruction is considerable, with 15.4% of housing on average

[18]Details of how we aggregated data on individuals to the household level are given in Appendix A.3.

destroyed in regions with low destruction and 49.3% in regions with high destruction (row 4). Moreover, regions which are closer to the inner-German border tended to be less destroyed than those that are further away (row 6). The pattern in income per capita is interesting: while regions with lower wartime destruction are slightly poorer in 1985 and 1989, they are slightly richer than the average region in 1995.

Panel B of Table 1 presents the data on West German firms in 2007, split up by regions above and below the median level of wartime destruction. On average, firms in regions with lower wartime destruction are slightly smaller as measured by the number of subsidiaries they operate in West Germany (row 1). Nevertheless they are also more likely to operate a subsidiary or branch in the East (8.3% versus 7.2%). On average, 7.7% of the firms in our sample operate a subsidiary or a branch in East Germany (row 3) and 1.8% operate in non-EU countries.

1.4 Social Ties and Regional Economic Growth

We first explore the effect of social ties between West and East Germans on income growth in West German regions. The structural equation of interest is

$$\log\left(\frac{y_{r,t}}{y_{r,1989}}\right) = \beta s_{r,1989} + \phi \log y_{r,1989} + Z_r'\zeta + \varepsilon_r \qquad (1)$$

where $y_{r,t}$ is income per capita in region r in year t, $t \in \{1991, 1993, 1995, 1997, 1999, 2001\}$. The left hand side variable is thus the growth in income per capita between 1989 and subsequent census years. $s_{r,1989}$ denotes our measure of social ties in region r. Z_r is a vector of controls, which always contains a constant term, a complete set of federal state fixed effects, and the distance between district r and the inner-German border. The coefficient of interest is β, which measures the effect of social ties on growth in income per capita after 1989. In all specifications we control for income per capita in 1989. The coefficient ϕ thus measures the degree of mean reversion in income per capita between West German districts. In our standard specification we also control for the pre-existing growth trend by including the log of growth between 1985 and 1989, $\log(y_{r,1989}/y_{r,1985})$. In these specifications the coefficient β thus estimates the *differential change* in the growth rate of income per capita after 1989 for regions with different levels of intensity of social ties to the East. The assumption that the relationship between growth in income per capita and social ties is linear is made for simplicity. The error term ε_r captures all omitted influences, including any deviations from linearity. Throughout, standard errors are calculated using the Huber-White correction to ensure robustness against arbitrary heteroscedasticity.

Equation (1) will consistently estimate the parameter of interest if $Cov(s_{r,1989}, \varepsilon_r) = 0$. This covariance restriction may not, however, hold in practice, since the settlement of expellees in West Germany prior to 1961 (and thus the strength of social ties to East Germany) may be correlated with persistent differences in growth prospects across regions. Although we show results of Ordinary Least Squares estimates of equation (1) for reference and comparison, we primarily rely on an instrumental variables strategy, which uses only the variation in $s_{r,1989}$ that is attributable to variation in wartime destruction across regions in 1946. Our first-stage specification is

$$s_{r,1989} = \gamma w_r + \phi^{fs} \log y_{r,1989} + Z_r' \zeta^{fs} + \nu_r, \tag{2}$$

where w_r is our measure of wartime destruction and (2) contains the same covariates as (1). Our key identifying assumption is that $Cov(w_r, \varepsilon_r) = 0$. It states that, conditional on the covariates we control for, (i) wartime destruction in 1946 has no effect on changes in growth in income per capita after 1989 other than through the settlement of expellees via the Soviet sector and (ii) there is no omitted variable which drives both wartime destruction and differential changes in income growth post-1989.

1.4.1 The First-Stage Relationship

Panel A of Table 2 shows our basic first-stage regressions, using the share of expellees via the Soviet sector in 1961 as a proxy for social ties in 1989. Column 1 is the most parsimonious specification as shown in equation (2). It regresses the share of expellees via the Soviet sector on the share of housing destroyed in 1946, while controlling for the distance to the inner-German border and for income per capita in 1989. The coefficient estimate of -0.019 (s.e.= 0.004) is statistically highly significant and suggests that a one standard deviation increase in the share of housing destroyed in 1945 (s.d.= 0.21) is associated with a 0.4 percentage point drop in the share of expellees via the Soviet sector in 1961. (This corresponds to 8% fewer expellees via the Soviet sector relative to the mean share of expellees via the Soviet sector across regions.)[19]

As expected, the share of expellees in 1961 falls with the distance to the inner-German border. The coefficient on income in 1989 is positive and significant, suggesting that expellees tended to settle in districts that were richer in 1989, which is most likely attributable to persistent differences in income per capita between regions which existed prior to 1961.[20]

[19]We do not have reliable data on the settlement patterns of refugees arriving from the Soviet sector during the same period. However, since both groups faced similar constraints regarding the shortage of housing, the settlement pattern of refugees across West German regions was likely very similar to that of expellees arriving from the Soviet sector. As both groups were roughly of the same size (around 3.5 million) we may speculate that a one standard deviation increase in wartime destruction may be associated with a drop in the total share of migrants from the East settling in a given West German region which is around twice as large.

[20]Income per capita in 1989 is included in all specifications to present the first stage corresponding to the instrumental variables results discussed below. If we drop all controls from the regression, the estimated coefficient is identical, -0.019, s.e.=0.007.

The specification in column 2 is our standard specification. It adds income growth in the five years prior to 1989 as an additional control. The coefficient of interest remains virtually unchanged at -0.020 (s.e.= 0.005). The coefficient on income growth is statistically indistinguishable from zero, suggesting that the pattern of settlement of expellees via the Soviet sector in 1961 is not correlated with income growth in the years prior to the fall of the Berlin Wall.

Figure 1 plots the conditional relationship estimated in this column and shows that the first-stage relationship is not driven by outliers. Columns 3-6 of Panel A of Table 2 show the first-stage regressions corresponding to robustness checks performed in the instrumental variables estimation. In column 3 we use the volume of rubble per capita in 1946 as an alternative measure of wartime destruction, which again yields a negative and significant coefficient. In column 4 we replace our control for the distance to the inner-German border with a fixed effect for each distance quartile and in column 5 we add the share of the workforce employed in agriculture, manufacturing, services, and government in 1989 (in both cases we do not report the coefficients on these variables for expositional clarity). Finally, column 6 adds the extent of migration after reunification as an additional control. In each case the coefficient of interest remains virtually unchanged and statistically significant at the 1% level.

Panel B of Table 2 repeats the same specifications as in Panel A, using the share of households with ties to relatives in East Germany in 1991 from our household level dataset as an alternative proxy for social ties in 1989. In the interest of preserving space we show only the coefficient of interest. All estimates are negative and all except the ones in column 3 and 4 are statistically significant at the 5% level (the latter is significant at 10%). The coefficient in column 1 is -0.099 (s.e.= 0.042). It implies that a one standard deviation rise in the share of housing destroyed in 1945 is associated with a 2.08 percentage point drop (or alternatively a 9.3% drop relative to the average) in the share of respondents that have a relative in East Germany in 1991. Similar results (not shown) hold for the share of respondents that report to be in contact with friends in East Germany. In the remainder of the paper we use the share of expellees via the Soviet sector in 1961 as our main proxy for social ties since, coming from a comprehensive census, we expect it to be measured with less error than the variables generated from our household level dataset. Needless to say, the correlation between the two proxies is very high (64%), as shown in Figure 2.

1.4.2 The Reduced Form Relationship

As a prelude to our instrumental variables estimates, Panel C shows the reduced form relationship between growth in income per capita after the fall of the Berlin Wall and wartime destruction. All specifications (except the one in column 3) are again identical to the ones in Panels A and B, with the left hand side variable now being the growth in income per capita between 1989 and 1995, $\log\left(\frac{y_{r,1995}}{y_{r,1989}}\right)$. The coefficient of interest is negative and statistically significant at the 5% level in all columns except in column 1, where it is significant at the 10% level. The estimate in column 2 is -0.048 (s.e.= 0.020), suggesting that regions which were least destroyed during the war experienced a significantly higher increase in the growth rate of income per capita post 1989 than regions which were most destroyed during the war. A one standard deviation drop in the share of housing destroyed in 1946 is associated with a 1.5 percentage point higher growth in income per capita over the six years following German reunification. The size of the estimated coefficient is stable across columns 1, 2 and 4-6, with point estimates ranging from -0.042 in column 1 to -0.052 in column 4. Figure 3 depicts this relationship graphically in a conditional scatter plot, where the slope shown corresponds to the estimate in column 2.[21]

As a first test of the mechanism by which wartime destruction could suddenly affect economic growth 50 years later, the specification in column 3 includes both the share of housing destroyed in 1946 and rubble per capita in 1946. The results are encouraging for our identification strategy: while the coefficient on the share of housing destroyed remains negative and significant at -0.060 (s.e.= 0.027), the coefficient on rubble per capita is positive and insignificant. This pattern suggests that it is primarily the lack of housing in 1946 and not wartime destruction per se that affects economic growth post 1989.

1.4.3 Instrumental Variables Results

In our instrumental variables estimation we explicitly test the main hypothesis of this paper: that the intensity of social ties between East and West Germans in 1989 is causally related to a rise in the growth rate of income per capita after the fall of the Berlin Wall. In Table 3, we estimate (1) using only the variation in the social ties in 1989 that is due to variation in wartime destruction, by instrumenting for the share of expellees via the Soviet sector in 1961. In column 1 we instrument with the share of housing destroyed in 1946. The coefficient estimate for the share of expellees is 2.169 (s.e.= 0.947), suggesting that a one standard deviation increase in the share of expellees in 1961 (s.d.= 0.019) is associated with a 4.3% rise in income per capita over the six years following 1989 (or roughly a 0.7 percentage point higher

[21] In the plot, Wilhelmshaven looks like a significant outlier. Dropping Wilhelmshaven from the sample reduces the coefficient estimate to -0.033 (s.e. = 0.016). As a more systematic check for the effect of outliers, we run a robust regression (according to the terminology used by STATA) in which observations with a Cook's D value of more than one are dropped and weights are iteratively calculated based on the residuals of a weighted least squares regression. The robust estimate is -0.032 (s.e. = 0.014).

rate of growth per annum).[22] The coefficient on income in the base year, 1989, is negative and significant, which suggests mean reversion in income per capita across West German districts. Somewhat surprisingly, the coefficient on the distance to the inner-German border is positive, which suggests that the districts closest to the inner-German border did not immediately profit from the opening of the border (which is in line with a similar observation in Redding and Sturm (2008), that the population of West German cities close to the inner-German border grew relatively little between 1989 and 2002).

Column 3 gives our standard specification in which we control both for the level of income in 1989 and for income growth in the five years preceding 1989. The coefficient of interest rises slightly to 2.442 (s.e.= 0.880) and is now significant at the 1% level.[23] The fact that we control for both for the pre-existing income level and for pre-1989 income growth means that this estimate is specific to the period after the fall of the Berlin wall: It can neither be explained by mean reversion in income growth nor by a pre-existing trend.[24]

The results of column 3 are almost unchanged when we simultaneously instrument the share of expellees with both the share of housing destroyed and with rubble per capita (shown in column 4). Column 2 shows the OLS estimate of our standard specification for comparison. It is only about one half of a standard error lower at 1.963 (s.e.= 0.574), suggesting that the endogenous assignment of expellees to West German districts induces only a relatively mild downward bias in the OLS estimate. A Hausman test fails to reject the null hypothesis that $Cov\left(s_{r,1989}, \varepsilon_r\right) = 0$.

1.4.4 Validity of the Exclusion Restriction

While the endogenous assignment of expellees to West-German districts does not seem to have a large impact on our results, our identifying assumption, that the degree of wartime destruction in 1945 affected growth in income per capita after 1989 only through its effect on the intensity of social ties in 1989, cannot be tested directly. Nevertheless, we can perform a number of falsification exercises to assess its plausibility. There are two types of potential challenges and corresponding tests.

[22]The F-statistic against the null that the excluded instrument is irrelevant in the first-stage regression is 22.56 (this is the squared t-statistic from Table 2).

[23]We do not cluster the standard errors at the federal state level as there are only 10 federal states. However, the results presented throughout this section are generally robust to doing so. For our standard specification the clustered standard error is 1.288, which implies that the coefficient of interest remains statistically significant at the 10% level.

[24]Since our model contains a lagged dependent variable there may be a mechanical bias in the coefficient of interest. In principle we can instrument for the lagged dependent variable as suggested by Anderson and Hsiao (1982). However, we do not have sufficient pre-1989 data to do this for the specification in column 3. However, we can use income in 1985 as instrument for income in 1989 in the specification of column 1. The coefficient estimate increases slightly to 2.289 (s.e.=0.984) and remains significant at the 5% level.

Simple Challenge The 'simple' challenge to our identifying assumption is that wartime destruction (or an omitted variable driving it) may have had a lasting effect on income growth in West German regions which persisted for more than half a century (until 1995). We believe that we can convincingly discard this 'simple' challenge.

First, our standard specification controls for the growth rate of income pre-1989 and thus identifies *changes* in the region-specific growth trajectory that occur after 1989.

Second, the conventional view in the literature is that wartime destruction had no lasting impact on West German growth post 1960 (Brakman, Garretsen, and Schramm (2004)). Figure 4 replicates part of this result. It shows coefficient estimates of regressions relating the population growth in West German cities to wartime destruction in 1946 for various years since 1929. Not surprisingly, wartime destruction had a strong and significant negative effect on population growth during the war (between 1939 and 1945). During the period of reconstruction between 1946 and 1960 the cities most heavily destroyed grew fastest. However, from 1960 onwards there is no statistically significant effect of wartime destruction on population growth and the coefficient estimates are virtually zero. To the extent that population growth is correlated with income growth, this result suggests that the effects of wartime destruction on income growth were short lasting.

Third, for the short period pre-1989 for which we have income data we find that growth in income per capita is uncorrelated with wartime destruction and with the settlement of expellees. Panel B of Table 3 shows a placebo experiment in which we use income growth between 1985 and 1989 as the dependent variable rather than as a control. All specifications are parallel to those in Panel A (except that we now control for log income per capita in 1985 rather than in 1989 and no growth trend, since we do not have the data). In all specifications in the panel the coefficient of interest is statistically indistinguishable from zero. Wartime destruction thus becomes relevant for economic growth only post 1989. In Table 14 we pinpoint the timing of the effect at a higher frequency by regressing log income per capita for each district and year post-1985 on the interaction of year fixed effects with the share of expellees in 1961. The table shows no effect of the settlement of expellees on income growth rates prior to 1989, a positive effect in all years post 1989 and a statistically significant effect on growth between 1989 and 1993 and later years. The timing of the effect is thus highly supportive of the view that variation in the degree of wartime destruction only became relevant after the fall of the Berlin Wall.

Sophisticated Challenge The more 'sophisticated' challenge to our identifying assumption is that the pattern of wartime destruction (or some omitted variable driving it) affected income growth through some other channel which only switched on post-1989.

One such possibility is that the allies may have bombed areas that were highly industrialized and the manufacturing sector may have experienced a relative decline in those areas after 1989. To address this potential concern, the specification in Table 3 column 5 controls for the sectoral composition of the workforce in 1989. Indeed, the estimated coefficient on the share of the workforce employed in manufacturing is negative, but it is not statistically significant and the coefficient of interest rises only marginally to 2.772 (s.e.=0.854). Any variation in income growth post 1989 due to the relative decline of manufacturing is thus unrelated to the variation in income growth post-1989 due to the settlement of expellees in West Germany. In addition, Figure 4 shows that wartime destruction was uncorrelated with pre-war population growth. To the extent that we can take pre-war population growth as an indicator for economic growth more generally, this would suggest that allied bombings during World War II were not specifically targeted at destroying cities which were on a higher or lower growth trajectory.

Another potential concern is that after 1989 highly skilled workers from East Germany may have migrated to the same districts in which their relatives settled before 1961, and that this migration may have increased the average wage paid in these districts. In column 6 we control for the flow of migration from East to West, and again there is little effect on the coefficient of interest.[25]

While neither of these two channels appear to be driving our results, there might be other omitted variables which are correlated with the pattern of wartime destruction and post-1989 deviations from the existing growth trajectory. Alternatively, we may be misinterpreting our results in that expellees may affect income growth after 1989 through some channel other than social ties, which switches on post-1989. In particular, expellees might have been somehow different from other Germans, and these different traits (a higher propensity to become an entrepreneur, different preferences, etc.), may have put them in an advantageous position to earn higher incomes post 1989 for reasons unrelated to social ties to the East. We are able to provide evidence on this, and the entire class of 'sophisticated' challenges, by comparing the expellees via the Soviet sector with expellees who arrived directly from the annexed parts of pre-war Germany. The direct expellees migrated from the same areas in Eastern Europe, they look very similar to expellees via the Soviet sector on observable characteristics, and their settlement pattern was affected by wartime destruction in a similar way. The only relevant difference between the two groups is that expellees who arrived directly from the annexed areas did not spend any significant time living (and forming social ties) in East Germany. If we misinterpret our results and the effects we document are driven by some omitted

[25]When we use the flow of migration from East Germany as the dependent variable, the coefficient on expellees is not significant, which is comforting for our interpretation of the results. A related result in the literature is that high-skilled workers from the East were actually less likely to migrate to the West than low-skilled workers up until 1996 (Fuchs-Schündeln and Schündeln (2009)).

variable which determined both wartime destruction and post 1989 income growth, or if there was something special about expellees per se that gave them access to business opportunities post-1989, we would expect to find the same effects for both the expellees via the Soviet sector and the direct expellees.[26]

As a prelude to our placebo test, Table 15A compares both groups of expellees using data from the 1971 census, the last census in which the two groups are separately identified. The table gives the average income, the average number of years of schooling, and the occupational structure of both groups of expellees and of the 'native' West German population. While both groups of expellees look different from native West Germans (they are slightly poorer, slightly less educated, and significantly less likely to be entrepreneurs), they look remarkably similar to each other. Expellees via the soviet sector have an average income of DM 777.8 and on average 9.69 years of schooling, direct expellees an average income of DM 764.4 and on average 9.63 years of schooling; Both groups have an identical share of entrepreneurs among them (3%) and a very similar occupational structure. Expellees via the Soviet Sector and direct expellees are thus extremely similar on observable characteristics.

In Appendix Table 15B we show that the same is true for the 1989 characteristics of the regions in which the two groups of expellees settle. The table reports regressions of district characteristics in 1989 on the share of both groups of expellees in 1961 and our standard region level controls (distance to east, income in 1989, income growth between 1985 and 1989, and state fixed effects). Each line of the table corresponds to one regression. It reports the dependent variable, the coefficients on the two groups of expellees, as well as the p-value corresponding to the null hypothesis that the two coefficients are equal. We cannot reject this hypothesis at the 5% level in any of the seven specifications. However, it appears that regions with a larger share of direct expellees also tend to have a somewhat larger share of the workforce emplyed in the agricultural sector (here the p-value is 5.9%). With this possible caveat, there do not appear to be any systematic differences between the regions in which the two groups settle.

In our placebo experiment we relate growth in income per capita post 1989 simultaneously to the share of expellees via the Soviet Sector and to the share of direct expellees, again conditinal our standard region level controls. The results are in Table 4. Column 1 gives the results from an OLS regression. While the coefficient on expellees via the Soviet Sector is positive, statistically significant at the 1% level, and very similar to the estimates from Table 3 (2.131, s.e.=0.706), the coefficient on the share of direct expellees is negative and statistically indistinguishable from

[26]Recall from our discussion in section 1.2 that expellees were allocated to the four sectors according to quotas and that, as far as we can tell from the statistics published in East Germany, the vast majority of expellees who were allocated to the Soviet sector migrated to the West before 1961.

zero. In columns 2 and 3 we add additional controls for the share of the population in agriculture (column 2) and for the share or the population employed in the other three sectors (column 3), which makes little difference to the results.

In columns 4-6 we repeat this exercize using our instrumental variables strategy. To compare the causal effect of direct expellees and expellees via the Soviet sector on income growth post 1989 we require two instruments which give us differential leverage in identifying the exogenous components in the settlement patterns of both groups. In Panels B and C we re-run our standard first-stage regression from Table 2 column 2, but include both the share of housing destroyed and the volume of rubble per capita in 1946. (We again do not report covariates in the interest of space.) Panel B gives the results for expellees arriving via the Soviet sector and Panel C gives the results for expellees arriving directly in West Germany. In the case of the former, the share of housing destroyed is significant with a negative sign and rubble is insignificant across all three specifications. In the case of the latter, the size of the effect of the share of housing destroyed is roughly preserved, though it is less precisely estimated. Importantly for us, the coefficient on the amount of rubble is negative and significant. Our two measures of wartime destruction thus give us differential leverage in identifying the exogenous components in the settlement patterns of both groups. We believe that this feature of the data is related to the timing of the arrival of the two groups of expellees. The direct expellees arrived immediately after the war, whereas the expellees via the Soviet sector arrived between 1945 and 1961. We therefore suspect that rubble per capita measures a dimension of wartime damage which was more important in the immediate aftermath of the war but was then cleared away relatively quickly, while the destruction of the housing stock had longer-lasting effects.[27]

Using both instruments, we are thus able to separately estimate the causal effects of expellees via the Soviet sector and of direct expellees on differential income growth after 1989. Columns 1-3 of Panel A present the results. While the coefficient on the share of expellees who arrived via the Soviet sector is again positive, similar in magnitude to the estimates obtained earlier (3.422, s.e.= 1.809 in column 1), and statistically significant at the 10% level, the coefficient on the share of direct expellees is close to zero and statistically insignificant. The growth effects we document are thus particular to expellees who had lived in the Soviet Sector and thus particular to the group which uniquely had the opportunity to form social to East Germans before moving to the West. As the two groups of expellees look extremely similar in other dimensions, we view these results as strong support in favor of our interpretation.

[27]Many German cities famously have 'rubble mountains' which were piled up in the first two to three years after the war.

1.4.5 Remaining Caveats

A remaining caveat to our results is that they are also consistent with two proximate interpretations which we do not emphasize due to our reading of German history. First, the results might document the value of 'local knowledge' about East Germany. In principle, the effects we document could be driven by individuals who lived in East Germany during their youth, do not have personal contact with anyone there, but remember enough details about the local economy to earn rents after reunification. We do not emphasize this interpretation as the conventional view is that economic conditions in East Germany have changed dramatically between 1961 and 1989, so that local knowledge acquired before the division of Germany would be useless after 40 years of socialist rule. Moreover, our individual-level results show that even households headed by individuals who were too young to remember living in East Germany experience a rise in their personal income after 1989 if they have a relative in the East, providing some partial evidence against this alternative interpretation.

Second, some of the rise in household and regional income may be driven by restitutions and payments of compensation to those whose property had been expropriated in East Germany. Under the reunification treaty, former owners of firms and real estate could apply for restitution providing that they had not received compensation from the East German government and the assets they were claiming still existed at the time of filing a claim. This meant that practically all individual claims made related to buildings and land. While there were a large number of claims filed (around 2 million), we do not believe that restitution or compensation payments could be responsible for the patterns we document in the data.

The administrative backlog created by the task of tracing ownership rights over a period of 50 years was so enormous that it was not sufficiently cleared in time to confound the effects we document (in fact, the authorities and courts are still processing claims to the present day). The first compensation payments were not set to begin until 1996, whereas the effects we document begin immediately after 1989 (Southern (1993)). The only real concern is thus the existence of cases in which the filing of a claim resulted in the restitution of a property prior to 1995, and the restitution of this property resulted in a rise in income. The early restitutions were mainly of firms, as the establishment of safe property rights for productive assets was a political priority (although even these were the subject of protracted legal battles, so that in 1992, 90% of claims concerning firms were still on hold due to litigation; see Sinn and Sinn (1992)).

We address this concern by calculating an upper bound for the value of all restitutions that could possibly have been made in the period in question and comparing it to the magnitude of the effects that we attribute to social ties. According to the government agency handling restitutions, half of all approved claims had been settled by restitution, and the total sum of compensation payments made between 1990 and 2009 was EUR 1.4 bn.[28] These compensation payments were made at about 50% of market value, and so a reasonable estimate of the value of all restitutions is therefore $\frac{EUR\,1.4bn}{0.5}$ =EUR 2.8bn. Assuming that all restitutions had been completed before 1995 (which they were not) and assuming that the new owners immediately sold the returned assets on (which they were often legally prevented from doing), aggregate income in West Germany could thus have risen at most by EUR 2.8bn. This number is an order of magnitude smaller than the total effect implied by our regional level results (around EUR 74.0 bn in 1989 equivalents of Euros).

1.5 Understanding the Effect on Regional Economic Growth

1.5.1 Entrepreneurial Activity

In an effort to shed light on the channel linking social ties to regional economic growth, we disaggregate regional income per capita into the average income of households whose primary income derives from entrepreneurial activity (entrepreneurs) and the average income of all other individuals (non-entrepreneurs) for each year.[29] In columns 1 and 2 of Table 5 we re-run our standard specification from column 3 in Table 3 with the growth rate in the average income of entrepreneurs and non-entrepreneurs as the dependent variables. Both specifications include the same covariates as our standard specification, but add the (log of the) average income of entrepreneurs and non-entrepreneurs in 1989, respectively, as an additional control. The coefficient estimate is 4.008 (s.e.= 2.096) for entrepreneurs (column 1) and 1.755 (s.e.= 0.735) for non-entrepreneurs (column 2); suggesting that a one standard deviation rise in the share of expellees via the Soviet sector is associated with a 7.6% rise in the average income of entrepreneurs, but only a 3.3% rise in the average income of non-entrepreneurs. Entrepreneurs who lived in a region with strong social ties to the East thus experienced a much steeper rise in their average income than non-entrepreneurs living in the same region.[30]

This strong effect on the income of entrepreneurs is mirrored by an increase in the number of entrepreneurs. In column 3 we re-run our standard specification but use

[28] Personal correspondance with Dr. Händler, press liaison of the Bundesamt fuer zentrale Dienste und offene Vermögensfragen.

[29] In the German Mikrozensus these are households whose household heads declare that their primary occupation is 'entrepreneur' (Selbststaendiger mit oder ohne Beschaeftigte).

[30] In order to determine whether these two coefficients are significantly different from each other we re-estimate the two equations as well as the first stage jointly with a 3SLS estimator. In this case the coefficient of interest in column 1 rises to 4.920 (s.e.= 1.657) and is significant at the 1% level. In column 2 the coefficient of interest is estimated to be 1.451 (s.e.= 0.665) and significant at the 5% level. In these specifications we control for the (log of the) average income of entrepreneurs and non-entrepreneurs in 1989 in both structural equations of interest and the first stage. The null of equality of these coefficients is rejected at the 5% level.

the share of the population that are entrepreneurs in 1995 as the dependent variable, where we again add the share of the population that are entrepreneurs in 1989 as an additional control. The coefficient of interest is 0.322 (s.e.=0.163), suggesting that a one standard-deviation rise in the share of expellees in 1961 (0.019) induces a 0.61 percentage point rise in the share of the population which are entrepreneurs. This is a sizable effect, corresponding to a 14.2% rise relative to the mean share of entrepreneurs in 1989 (0.043).

1.5.2 Firm Investment

The strong rise in entrepreneurial income in regions with strong social ties to the East suggests that firms which were based in these regions generated higher profits in the years following the fall of the Berlin Wall. One possible reason for such a rise in profitability is that locating in a region with strong social ties to the East may have generated a comparative advantage in investing in the East. Firms who had access to a workforce or to an owner with strong social ties to the East may have been in a better position to assess the value of East German firms that came up for sale or may have been better able to gage demand for Western products and services. We explore this possibility by examining the holdings of subsidiaries and branches of West German firms in the East.

Our firm level data is from 2007, which is the year in which the ORBIS dataset expands coverage to small and medium sized German firms. We have data on 19,402 firms whose headquarters are located in West Germany and who operate at least one subsidiary or branch. For these firms we calculate a dummy variable which is one if the firm operates a subsidiary or branch in East Germany and zero otherwise. Since West German firms could not own assets in East Germany prior to the fall of the Berlin Wall, any subsidiaries or branches that they operate in 2007 must have been acquired after 1989. Our dummy variable is thus informative both about the investment behavior of West German firms in East Germany since 1989 and about a possible long-lasting effect of social ties in 1989 on the economic structure of West Germany.

The structural equation of interest is

$$b_{kdr,2007} = \beta^f s_{dr,1989} + \phi^f \log y_{d,1989} + Z'_{kdr}\zeta^f + \varepsilon^f_{kdr} \qquad (3)$$

where $b_{kdr,2007}$ stands for the dummy indicating whether firm k in West German district d and region r operates a subsidiary or a branch in East Germany in 2007. $s_{dr,1989}$ is again our measure of social ties between the residents of district d in region r and East Germany in 1989; $y_{r,1989}$ stands for income per capita in region r in 1989; and Z_{kdr} is a vector of firm and district level controls which contains a complete set of federal state fixed effects, a fixed effect for the sector in which the firm has its primary operations, the log of the number of subsidiaries and branches that firm k operates in West Germany, and the distance between district d and the inner-German border. Note that income per capita in 1989 is available only at the regional level and not at the district level.

The coefficient of interest is β^f which measures the effect of the intensity of social ties to the East in a given West German district in 1989 on the probability that a firm headquartered within that district operates a subsidiary or branch in East Germany in 2007. As in section 1.4, we account for the possibility that our measure of social ties (the settlement of expellees via the Soviet sector in West Germany) is jointly determined with income growth by instrumenting $s_{dr,1989}$ with the share of housing destroyed in 1946. The first stage of our instrumental variables strategy is thus the analog to (2). We cluster all standard errors at the district level to account for likely spatial correlation.

Panel A of Table 6 shows reduced form estimates, relating the share of housing destroyed in 1946 directly to the probability that a given firm operates a subsidiary or branch in East Germany in 2007. In column 1, we regress our dummy variable on the share of housing destroyed in the district and the log of the number of subsidiaries and branches that the firm operates in West Germany in 2007, which we use as a simple control for the size of the firm.[31] The coefficient of interest is -0.030 (s.e.=0.011) and statistically significant at the 1% level. The estimate implies that a one standard deviation rise in the extent of wartime destruction within a given West German district is associated with a 0.7% drop in the probability that a firm based in that district operates a subsidiary or branch in East Germany in 2007.[32] Unsurprisingly, the coefficient on our size control is positive and significant, refecting the fact that larger firms also are more likely to operate in East Germany. Columns 3-6 add all of the now familiar district and region level covariates from section 1.4, and column 3 gives the analog of our standard specification. Throughout, the coefficient of interest remains in a tight range between -0.029 and -0.031 and is statistically significant at 1%.

Panel B shows our instrumental variables estimates of equation (3), which use the variation in wartime destruction to quantify the causal effect of social ties in 1989 on the investment behavior of West German firms. All specifications contain the same covariates as those in Panel A. The estimates in all columns are positive and all are statistically significant at the 5% level. The estimate from our standard specification in column 3 is 1.556 (s.e.= 0.693), which implies that a one standard deviation rise in the share of expellees in a West German district is associated with a 3.4% increase in the probability that a firm based in that district will operate a

[31]The raw correlation between our dummy variable and the share of housing destroyed is -0.019 (s.e.=0.012) and statistically significant at the 10% level.

[32]The standard deviation of the share of housing destroyed at the district level is 0.24.

subsidiary or a branch in East Germany in 2007.[33] This link between the pattern of social ties in 1989 and the investment behavior of West German firms offers a potential explanation for why residents of districts with strong social ties to the East, and entrepreneurs in particular, may have experienced a rise in their incomes. The increased firm level activity also suggests potential spill-over effects of social ties, i.e. residents of districts with strong social ties to the East might experience a rise in income even if they themselves did not have social ties to the East. We return to this issue in the following section where we calculate the size of such potential spill-over effects.

The remaining panels of Table 6 show the results of a number of falsification exercises. If the pattern in holdings of subsidiaries and branches prevailing in 2007 is truly attributable to the intensity of social ties to East Germany in 1989, and not to some other factor, our measure of social ties to East Germany should predict investment in East Germany but not in other areas of the world.[34] Panels C-F repeat the same specifications as in Panel B, but with a dummy variable indicating whether a firm has subsidiaries or branches in Poland, in the 'old' EU countries (the 14 member countries other than Germany prior to the enlargement in 2004), in the 'new' EU countries (the 9 countries, other than Poland, which joined the EU in 2004), or in non-EU countries as the dependent variable. As expected, all estimated coefficients in Panels D, E and F are statistically indistinguishable from zero. Firms which are based in districts with a high share of expellees are thus *not* more likely to operate subsidiaries or branches in areas other than East Germany. Interestingly, however, the only exception from this rule is that the estimates for Poland are positive and statistically significant at the 5% level in all columns. The estimated effect for Poland is about 1/5th the size of the effect estimated for East Germany. Since the largest group of expellees who settled in West Germany after 1945 actually came from areas that are today part of Poland, these results suggest a possible additional effect of social ties to Poland on the investment behavior of West German firms.[35] [36]

[33]The district level standard deviation of the share of expellees via the Soviet sector is 0.022.

[34]If firms from districts with a high fraction of expellees were merely good at capitalizing on new business opportunities, regardless of social ties, we might, for example, expect to see an effect on their holdings in other Eastern European countries following consecutive rounds of EU enlargement.

[35]However, the size of the coefficient for Poland is similar to that of some of the other, insignificant coefficients.

[36]Table 16 in the appendix reports the outcomes of additional robustness checks, in which we repeat the same specifications as in panels A and B of Table 6, but now use the share of each firm's subsidiaries and branches operated in East Germany as the dependent variable. The results are again similar, indicating that firms which are headquartered in West German districts that have strong social ties to East Germany in 1989 also operate a larger share of their subsidiaries and branches in East Germany today.

1.6 Social Ties and Household Income

Lastly, we explore how individual households may have profited from having ties to East Germany and how these ties may have affected regional economic growth. We use the German Socio-Economic Panel (SOEP) dataset and select households which were in the panel in 1985, located in West Germany in 1989 and remained in the panel at least until 1995. The 1991 wave of the panel contains several questions about contacts with friends and family in East Germany. Since the survey was conducted in the second year after the fall of the Berlin Wall and individuals had some time to renew ties with individuals in the other part of Germany, we choose not to rely on information about the intensity of contact to friends and relatives, although it is available.[37] Instead, we base our work on the response to the simple factual question: "Do you have relatives in East Germany?", and generate a dummy variable that is one if at least one member of the household responded with 'yes' and zero otherwise. A possible source of measurement error is that some West Germans may have migrated to East Germany directly after the fall of the Berlin Wall and before the conclusion of the 1991 wave of the survey. However, our data indicate that the flow of migrants from West to East in 1990 was not large enough to plausibly induce a quantitatively large bias. It is thus safe to assume that households which were based in West Germany in 1989 and report a relative in East Germany in 1991, also had a relative in East Germany in 1989.

Table 8 gives summary statistics for the entire panel of 1911 households, and for the subsets of households which report and do not report any ties to relatives in East Germany. The households with ties to East Germany had slightly lower income in 1989 (DM 3219 versus DM 3491 per month in 1989) and the household heads of households with ties to East Germany tend to have had slightly less education (on average 12.12 years versus 12.42 years). However, the two subsets of households look very similar on other observable dimensions: they had about the same amount of capital income on average, and the proportion of households whose head was reported as entrepreneur was similar across the two subsets, as was the proportion reported as unemployed.

Our basic household level regression estimates the growth in household income that results from having social ties to East Germany in 1989:

$$\log\left(\frac{y_{i,t}}{y_{i,1989}}\right) = \beta^{hh}T_i + \phi^{hh}\log y_{i,1989} + Z_i'\zeta^{hh} + \varepsilon_i, \qquad (4)$$

where $y_{i,t}$ is the income of household i in year t, T_i is a dummy variable indicating ties to East Germany and Z_i is a vector of controls which contains a full set of district

[37] Our results are very similar if we use this information on friendships or condition on respondents indicating 'close' ties to their relatives or friends.

fixed effects, household income growth between 1985 and 1989, and the gender, age and age squared of the household head. The coefficient of interest is β^{hh} which estimates the effect of ties to relatives in East Germany on the growth in household income after 1989.

Before estimating (4), Table 9 establishes the consistency of our household level and region level datasets. Column 1 reproduces the standard specification from Table 3 column 3, in which we related region-level income growth to the share of expellees in the district in 1961, regional average income in 1989, regional average income growth prior to 1989, and the distance of the region to the inner-German border. In column 2, we regress *household level* income growth on the same region level covariates, as well as on household level income in the base year and household level income growth between 1985 and 1989. The coefficient estimate on share expellees is insignificant, but remarkably similar to the one we obtained in the district level dataset, 2.777 (s.e.= 3.609). The fact that the estimate is statistically insignificant is not surprising as we are now using a sample of 1911 households, rather than district averages of 1% of the population. However, it is comforting that both datasets seem to have similar quantitative implications in this regard.

In column 3 we drop the region level share of expellees variable and replace it with a household level dummy variable indicating ties to East Germany. The estimate is positive and highly statistically significant, 0.069 (s.e.= 0.025). However, our dummy variable indicating ties to East Germany is mechanically correlated with the average level of social ties in a given district. The coefficient estimate in column 3 may therefore confound the direct (household) effect of having ties to the East with possible spill-over effects from other households in the same region having social ties.

We therefore introduce a full set of region fixed effects in column 5, which absorb the region wide effect of a higher incidence of social ties. Additionally we control for gender, age and age squared of the household head (which are standard controls in the labor literature) and estimate the full model in (4). The coefficient of interest drops to 0.049 (s.e.= 0.023) and remains statistically significant at the 5% level. The point estimate indicates that households with ties to East Germany in 1989 experienced on average a 5 percentage points higher income growth in the 6 years following the fall of the Berlin Wall relative to comparable households without such ties.

For specification (4) to consistently estimate the coefficient of interest we require $cov(T_i, \varepsilon_i) = 0$. As social ties existing prior to 1989 must have been maintained for reasons unrelated to any expected economic benefit, this condition does not fail due to reverse causality. However, it may still fail if individuals with social ties have some omitted characteristics which affect income *differently* after 1989 as compared to before 1989 and are correlated with social ties. In particular, individuals with social ties might somehow be more entrepreneurial or better educated, and therefore better able to seize the economic opportunities that present themselves after the fall of the Berlin Wall.[38] The summary statistics in Table 8 suggest the opposite: the heads of households with social ties to the East were actually slightly less educated and were somewhat less likely to be an entrepreneur in 1989. We control for these observable characteristics in our robustness checks discussed below.

In Table 10 we add a number of additional covariates (again this table only reports the coefficient of interest and the coefficients on the variables that are added relative to the standard specification). Column 1 reproduces the standard specification for comparison. Column 2 adds years of education and years of education squared of the household head in 1989, both of which remain statistically insignificant. Column 3 adds the log of capital income in 1989 as a proxy for the household's ability to finance investments. The variable is positive and significant, but induces little change in the coefficient of interest which remains at 0.047 (s.e.= 0.023). Column 4 introduces a dummy variable for whether the household head is an entrepreneur. Surprisingly, this variable remains insignificant, and again induces almost no change in the coefficient of interest. Finally, in column 5 we add a dummy for household heads who are not in employment in 1989 and column 6 adds all of these additional covariates simultaneously. Throughout, the changes in the coefficient of interest are minor and it remains statistically significant at the 10% level.

In Figure 5 we explore the timing of the effect by again using the full panel structure of the data in a specification analogous to Table 14 column 1.[39] We regress the income of each household in a given year on its income in 1985, a full set of year and district fixed effects, the interaction of year effects with the dummy variable indicating ties to East Germany in 1991, and controls from our standard specification (gender, age and age squared). The figure plots the interaction of year fixed effects with the dummy for ties to relatives in 1991, and identifies a 5% confidence interval.[40] These interactions measure the differential income growth of households with social ties to the East between 1985 and the indicated year. The pattern is striking: the estimates are statistically indistinguishable from zero until 1989, when the estimates jump up and are mostly statistically significant at the 5%

[38]A second reason why the estimation might be inconsistent is that we include a lagged dependent variable, which might be correlated with an auto-correlated error. We cannot perform Arellano-Bond style estimation as we do not have enough pre-1989 data. However, when running a specification in which we control for the household income in 1989, but not for the pre-existing growth trend, we can instrument for household income in 1989 with household income 1985. When running a specification equivalent to column 5 in Table 9, but without controlling for growth between 1985 and 1989, the coefficient of interest is estimated to be 0.049 (s.e.=0.023). When instrumenting for household income in 1989 with household income in 1985 the coefficient of interest is estimated to be 0.038 (s.e.=0.022) and significant at the 10% level.

[39]Although we use a standard fixed effects estimator as all of our covariates are exogenous

[40]The sample size is decreases monotonically from 1911 in 1995 to 1419 in 2001.

level until 1995. All coefficients after 1989 are statistically significant at the 10% level until the end of the sample in 2001.

Lastly, we can check that our results are robust to potentially confounding restitutions of assets from the East. The claimants who were likely to have received their assets by 1995, the year on which we focus in our analysis, are those whose claim related to a firm. If we re-run our standard specification in Column 5 of Table 11, dropping all households that state not to own productive assets in 1989, but state to own productive assets in 1995, the coefficient drops only very slightly to 0.047 (s.e.= 0.023).[41] The household level results are thus robust to excluding the subgroup of observations that could plausibly have benefited from restitutions prior to 1995. This adds to the evidence presented in section 1.4.4 that restitutions are unlikely to drive our results.

1.6.1 Cohort Heterogeneity

Our sample contains only 596 households with ties to the East. Nevertheless, we are able to give some further evidence on why the income of individual households increased after 1989. In Table 11, we estimate our standard specification, but interact the dummy for ties to East Germany with a fixed effect for the age quartile of the household head (and naturally also add fixed effects for each age quartile on the right hand side). The coefficient estimate for the youngest age quartile (those aged below 40 in 1989) is positive and significant at the 10% level, 0.092 (s.e.= 0.051). This is particularly interesting as it indicates that the effect we estimate is about knowing people and not about knowing places: the household heads in this group were younger than 11 years old at the time when the Berlin Wall was built and thus could not have had much personal experience of living in East Germany. However, they could easily have kept in contact with their relatives in East Germany.

1.6.2 East German Households

We have presented evidence on the returns to social ties accruing in West Germany. Naturally, we are interested in the corresponding effects in East Germany. Due to a lack of data on the settlement of expellees and on outcome variables prior to 1990, we cannot replicate our region level results for East Germany. However, we can replicate part of our standard specification in column 5 of Table 9 for households in East Germany. In particular, Table 17 shows results for a regression relating log income of East German households in the years after German reunification to a dummy variable indicating relatives in West Germany in 1991 and all household level covariates of column 5 in Table 9 other than income in 1989 and pre-1989

[41] A total of 57 households in our sample acquired productive assets between 1989 and 1995, 17 of which had relatives in East Germany.

income growth. As we have no data on income before 1990 for East German households, this specification can only speak to differences in levels of income, rather than to differences in income growth. Nevertheless, we find that the estimate on the coefficient of interest is positive in all years between 1990 and 1995 and marginally significant in two of the six years (1992 and 1994). For example, the estimate for 1995 is 0.057 (s.e.=0.040) suggesting that East German households with ties to the West tend to have higher income than those without ties to the West.

1.6.3 Spill-over Effects

The household level effect we estimated in this section is economically large. However, it is not large enough to account for the size of the entire region level effect of social ties estimated in section 1.4. Since the district level standard deviation of the share of households with social ties to the East is 0.10 (see Table 1), a one standard deviation rise is thus associated with at most a 0.7 percentage point rise in income per capita at the region level (Table 9, column 3). To the extent that the variation in the level of social ties in 1991 is driven by variation in the share of expellees (recall that the correlation between our two measures of social ties is very high), a one standard deviation rise in the share of households with relatives in East Germany corresponds to the same variation as a one standard deviation rise in the share of expellees in a district. However, a one standard deviation increase in the share of expellees in 1961 was associated with a 4.3 percentage point rise in district level income. The two estimates suggest a spill-over effect by which households without ties profit from living in a district with strong ties to the East.

This spill-over is consistent with the view that firms which were based in regions with strong social ties had an advantage in investing in East Germany, which resulted in increased wage levels in those West German districts. The presence of spillovers might also imply that social ties have an economic value even at the second or third degree of separation, where a household (and entrepreneurs in particular) might profit from having a social tie to another household who has a social tie to East Germany.[42]

1.7 Understanding the Microeconomic Effects

While we were able to provide a range of evidence suggesting that social ties facilitated regional economic growth by generating a rise in entrepreneurial activity, a deeper question is why social ties matter at the individual level. Two potential channels linking social ties to economic performance at the microeconomic level are that (i) social ties may reduce informational asymmetries (Granovetter (1973), Var-

[42] A novel by Schulze (2005) presents an example of such interaction at the second degree of separation.

ian (1990), Stiglitz (1990)) and (ii) social ties may sustain a large range of credit contracts that would otherwise not be feasible (Besley and Coate (1995)). In our context, possible examples of these stories might be (i) a West German who hears from his East German cousin about the lack of second hand cars in East Germany and reacts by starting a local car dealership in the East or (ii) a wealthy West German who lends his own money or acts as a guarantor for a East German relative who otherwise would not have access to funds.

One way to gauge the relevance of the first mechanism is to test whether social ties are particularly beneficial for firms in sectors in which informational frictions are important. In Table 7 we return to the standard specification of our firm level analysis (Table 6, Panel B, column 3) and replace the share of expellees variable with its interaction with each of the four sectoral fixed effects included in the specification (instrumenting these variables with the interactions of the sectoral fixed effects and the share of housing destroyed in 1946). The estimated effects in the Primary, Government, and Manufacturing sectors are statistically insignificant, while the effect estimated for the Trade and Services sector is positive and statistically significant at the 5% level (0.962, s.e.= 0.449). We may interpret this as evidence that social ties were particularly important for firm investment in the Trade and Services sector, which is arguably the sector of the economy which is most susceptible to informational asymmetries and reliant on knowledge of local demand.[43]

We may evaluate the relevance of the second channel by testing whether the West German households who had the highest ability to fund or act as guarantors for projects in the East were also the ones to benefit most from their social ties to the East. Column 1 of Table 12 replicates the standard specification of our household level results from Table 9, column 5 (again reporting only the coefficient of interest). In columns 2-4 we add a dummy which is one if the household's capital income in 1989 is above the pth percentile (75th, 95th and 99th, respectively) as well as the interaction of this variable with our dummy for social ties. If the principal function of social ties was to enhance households' ability to enter into credit contracts we would expect the coefficient on the interaction term to be positive. In column 2 the point estimate on our social ties dummy drops marginally to 0.46 (s.e.=0.026) and remains significant at the 10% level. The dummy indicating top-quartile capital income in 1989 is estimated to have a positive, but insignificant significant effect on post reunification growth in household income (0.046, s.e.= 0.028). However, the interaction of social ties and the above-median wealth dummy is insignificant. Columns 3 and 4 show similar results, using the 95th and 99th percentile. This evidence suggests that West German households benefited from their ties to East Germans regardless of their wealth in 1989 and, consequently, that social ties may

[43]It would be interesting to disaggregate this estimate for a finer partition of the trade and services sector. Unfortunately this is not possible with the data available.

have acted mainly as conduits for information rather than a form of 'social' collateral which would serve to sustain credit contracts.

1.8 Discussion

In this paper we used evidence from German reunification to establish a causal link between social ties and regional economic development. West German regions which, for idiosyncratic reasons, have strong social ties to East Germany at the time of the fall of the Berlin Wall exhibit substantially higher growth in income per capita in the early 1990s. This effect on regional economic growth appears to be driven both by a rise in the profits accruing to entrepreneurs and by an increase in the share of the population engaging in entrepreneurial activity. Consistent with these findings, we show that firms headquartered in a West German district which has strong ties to East Germany in 1989 are more likely to operate a subsidiary or a branch in East Germany today, suggesting that the pattern of social ties which existed in 1989 may thus have had a lasting effect on the pattern of economic activity in West Germany. We also show that West German households who have a relative in East Germany in 1989 experience a persistent rise in their personal income after the fall of the Berlin Wall. While this rise is economically large, it accounts for only about one sixth of the rise in income per capita at the region level, suggesting positive spill-over effects, through which households living in regions with strong ties to the East experience rising incomes even if they themselves do not have direct personal ties to the East.

Our findings appear robust to a wide range of plausible variations in the estimation strategy and placebo treatments. However, while we interpret these results as implying that social ties affect regional economic development, this interpretation is subject to two important caveats. First, we cannot rule out the possibility that social ties merely served to re-distribute rents between regions with stronger and weaker social ties to the East, rather than increasing economic output at the country level. However, since our results suggest that there were positive spill-overs from households with social ties to those households without such ties at the regional level, it is unclear why we should expect crowding out at higher levels of aggregation. Second, it is unclear how our findings generalize beyond the historical context of a large economic transition, such as the economic re-integration of Germany. While our results may be informative about the economic effects of social ties maintained by migrants and diasporas in other parts of the world, social ties might be particularly useful in an environment in which markets are established rapidly and informational asymmetries are large.

A Appendix

A.1 Tables

TABLE 1: SUMMARY STATISTICS (REGION AND FIRM LEVEL DATA)

	(1)	(2)	(3)
PANEL A: Region Level Data	*All*	*Low Destr.*	*High Destr.*
Share Expellees (Soviet sector) '61	0.048	0.049	0.047
	(0.019)	(0.022)	(0.015)
Share Expellees (Direct) '61	0.119	0.143	0.095
	(0.045)	(0.041)	(0.036)
Share Ties to Relatives '91	0.223	0.235	0.211
	(0.100)	(0.114)	(0.085)
Share Housing Destroyed '46	0.321	0.154	0.493
	(0.210)	(0.108)	(0.141)
Rubble '46 (m^3 p.c.)	0.090	0.037	0.144
	(0.070)	(0.033)	(0.056)
Distance to East (100km)	1.753	1.504	2.010
	(1.075)	(1.071)	(1.046)
Income 1985 (DM, p.c.)	1598	1568	1628
	(126)	(140)	(104)
Income 1989 (DM, p.c.)	1761	1747	1775
	(131)	(147)	(114)
Income 1995 (DM, p.c.)	2222	2227	2218
	(154)	(166)	(143)
Migration from East '91-'95	0.005	.006	0.004
	(0.004)	(0.005)	(0.002)
N	71	35	36
PANEL B: Firm Level Data	*All*	*Low Destr.*	*High Destr.*
S. & B. in West Germany (log)	0.443	0.437	0.450
	(0.742)	(0.729)	(0.756)
Share of Total S. & B. in East Germany	0.026	0.028	0.024
	(0.103)	(0.107)	(0.098)
S. & B. in East Germany (Dummy)	0.077	0.083	0.072
	(0.267)	(0.275)	(0.259)
Share of Total S. & B. in Non-EU Countries	0.006	0.006	0.006
	(0.055)	(0.055)	(0.054)
S. & B. in Non-EU Countries (Dummy)	0.018	0.018	0.017
	(0.131)	(0.134)	(0.129)
N	19420	9726	9694

Notes: The table presents means (and standard deviations). Variables in Panel A refer to our sample of regions used in Tables 2, 3, 4, 65, 14 and 15. Variables in Panel B refer to our sample of firms used in Tables 6, 7 and 16. Column 1 shows data for all observations. In Panel A, columns 2 and 3 show data for regions in which the share of housing stock destroyed was below the median and above median, respectively. In Panel B, columns 2 and 3 present means and standard deviations for firms headquartered in regions with share of housing stock destroyed above and below the median, respectively. Column 4 shows p-values of a t-test of equality of the means in column 2 and 3. The term 'S&B' stands for subsidiaries and branches which firms headquartered in a given West German region operate in the indicated location. Monetary values are given in nominal Deutsche Mark. See data appendix for details.

TABLE 2: WARTIME DESTRUCTION, SOCIAL TIES, AND INCOME GROWTH

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: First Stage	*Share Expellees (Sov. Sector) '61*					
Share Housing Destroyed '46	-0.019***	-0.020***		-0.021***	-0.021***	-0.020***
	(0.004)	(0.005)		(0.005)	(0.004)	(0.005)
Rubble '46 (m^3 p.c.)			-0.044***			
			(0.013)			
Distance to East (100km)	-0.005***	-0.005***	-0.005***		-0.004***	-0.005***
	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)
Income 1989 (p.c., log)	0.042***	0.047***	0.043***	0.046***	0.025*	0.047***
	(0.012)	(0.013)	(0.014)	(0.014)	(0.015)	(0.013)
Income '89/'85 (p.c., log)		-0.026	-0.020	-0.028	-0.029	-0.026
		(0.024)	(0.024)	(0.026)	(0.026)	(0.024)
Migration from East '91-'95						-0.006
						(0.212)
R^2	0.918	0.920	0.905	0.989	0.931	0.920
PANEL B: First Stage (altern.)	*Share Ties to Relatives '91*					
Share Housing Destroyed '46	-0.099**	-0.102**		-0.095*	-0.102**	-0.101**
	(0.042)	(0.043)		(0.050)	(0.041)	(0.047)
Rubble '46 (m^3 p.c.)			-0.161			
			(0.146)			
PANEL C: Reduced Form	*Income '95/'89 (p.c., log)*					
Share Housing Destroyed '46	-0.042*	-0.048**	-0.060**	-0.052**	-0.058***	-0.047**
	(0.021)	(0.020)	(0.027)	(0.020)	(0.020)	(0.020)
Rubble '46 (m^3 p.c.)			0.046			
			(0.071)			
N	71	71	71	71	71	71
Distance Dummies	no	no	no	yes	no	no
Sector Controls	no	no	no	no	yes	no

Notes: Coefficient estimates from Ordinary Least Squares regressions at the regional level. Standard errors are given in parentheses. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. The main variable of interest in all columns except column 3 is the share of the region's 1939 housing stock which was destroyed in 1946. In column 3 the main variable of interest in panel A and B is the amount of rubble per capita in the region in 1946. In column 3 of panel C both of these variables are included. The dependent variable in panel A is our main proxy for the intensity of social ties to East Germany - the share of the region's 1961 population which are expellees who had arrived from the Soviet sector. The dependent variable in panel B is our alternative measure of the intensity of social ties to the East: the share of the population which states in the 1991 SOEP survey to have relatives in East Germany. In Panel C the dependent variable is the log of the ratio of the region's mean per capita income in 1995 and 1989. All regressions include 10 federal state fixed effects. All specifications in panel B and C include the same controls as show in panel A. The coefficient estimates on these are suppressed for expositional clarity. The additional control in column 4 is the share of the population working in manufacturing in 1989. The additional control in column 5 is the number surveyed individuals who migrated to the region in the years 1991, 1993 and 1995 from East Germany as a share the total surveyed population (we do not have data for 1992 and 1994).

TABLE 3: SOCIAL TIES AND INCOME GROWTH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(IV)	(OLS)			(IV)		
PANEL A: Main Results					*Income '95/'89 (p.c., log)*		
Share Expellees (Sov. S.) '61	2.169**	1.963***	2.442***	2.453***	2.526***	2.772***	2.366***
	(0.947)	(0.574)	(0.880)	(0.877)	(0.885)	(0.854)	(0.878)
Distance to East (100km)	0.011**	0.008**	0.011**	0.011**		0.012***	0.011**
	(0.004)	(0.003)	(0.004)	(0.004)		(0.004)	(0.004)
Income 1989 (p.c., log)	-0.267***	-0.189***	-0.209***	-0.209***	-0.212***	-0.305***	-0.206***
	(0.068)	(0.060)	(0.060)	(0.060)	(0.063)	(0.072)	(0.062)
Income '89/'85 (p.c., log)		-0.362***	-0.355***	-0.355***	-0.379***	-0.278***	-0.353***
		(0.083)	(0.086)	(0.086)	(0.087)	(0.083)	(0.087)
Sh. Employed in Agricult. '89						-0.115	
						(0.295)	
Sh. Employed in Manufact. '89						-0.301	
						(0.283)	
Sh. Employed in Services '89						0.145	
						(0.290)	
Sh. Employed in Governm.'89						-0.522	
						(0.397)	
Migration from East '91-'95							0.349
							(1.130)
R^2	0.505	0.598	0.590	0.589	0.567	0.642	0.593
PANEL B: Placebo					*Income '89/'85 (p.c., log)*		
Share Expellees (Sov. S.) '61	-	0.656	0.560	0.557	0.672	0.443	0.790
	-	(0.602)	(1.024)	(1.029)	(0.939)	(1.100)	(1.041)
N	71	71	71	71	71	71	71
Distance Dummies	no	no	no	no	yes	no	no
Instruments	Housing	-	Housing	Housing & Rubble	Housing	Housing	Housing

Notes: The table reports coefficient estimates from instrumental variable regressions at the regional level in columns 1 and 3 through 7. Column 2 reports results from an Ordinary Least Squares regression. Standard errors are given in parentheses. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. In Panel A the dependent variable is in all regressions the log of the ratio of mean per capita income in 1995 and 1989. In Panel B it is the log of the ratio of mean per capita income in 1989 and 1985. The main variable of interest in all columns is the share of the region's 1961 population which are expellees who had arrived from the Soviet sector. In column 1, 3, 5, 6 and 7 we have instrumented for this with the share of the region's 1939 housing stock which was destroyed in 1946. In column 4 we use the amount of rubble per capita in 1946 as additional instrument. First stage results are shown in Table 2. All regressions include 10 federal state fixed effects. In Panel A all regressions control for the log of mean per capita income in 1989 and columns 2-7 include the log of the ratio of mean per capita income in 1989 and 1985 as control. In Panel B all regressions control for the log of per capita income in 1985. All regressions except column 5 control for a region's distance to the former East German border. Column 5 controls for 4 distance dummies, corresponding to the quartiles of the distance measure. The regression shown in column 6 controls for the share employed in four sectors in 1989 and the regression shown in column 7 controls for the number of regional immigrants from East Germany in the years 1991, 1993 and 1995 (we do not have data for 1992 and 1994). In Panel B we do not report results for these covariates for expositional clarity.

TABLE 4: PLACEBO

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A:	*Income '95/'89 (p.c., log)*					
		(OLS)			(IV)	
Share Expellees (Sov. S.) '61	2.131***	2.150***	2.039***	3.422*	3.396*	2.943*
	(0.706)	(0.727)	(0.561)	(1.809)	(1.787)	(1.738)
Share Expellees (Direct) '61	-0.092	-0.099	-0.043	-0.350	-0.371	-0.065
	(0.150)	(0.161)	(0.155)	(0.624)	(0.660)	(0.698)
Sh. Employed in Agricult. '89		0.047	-0.028		0.139	-0.114
		(0.193)	(0.304)		(0.272)	(0.297)
Sh. Employed in Manufact. '89			-0.197			-0.316
			(0.252)			(0.308)
Sh. Employed in Services '89			0.240			0.121
			(0.253)			(0.333)
Sh. Employed in Governm.'89			-0.452			-0.535
			(0.429)			(0.413)
R^2	0.600	0.600	0.664	0.557	0.561	0.640
Instruments	-	-	-	Housing & Rubble	Housing & Rubble	Housing & Rubble
PANEL B: First Stage				*Share Expellees (Sov. Sector) '61*		
Share Housing Destroyed '46				-0.020***	-0.021***	-0.020***
				(0.006)	(0.006)	(0.006)
Rubble '46 (m^3 p.c.)				0.002	0.001	-0.002
				(0.015)	(0.015)	(0.016)
PANEL C: First Stage				*Share Expellees (Direct) '61*		
Share Housing Destroyed '46				-0.026	-0.026	-0.027
				(0.018)	(0.018)	(0.020)
Rubble '46 (m^3 p.c.)				-0.107**	-0.106**	-0.104**
				(0.046)	(0.045)	(0.051)
N	71	71	71	71	71	71
Standard Controls	yes	yes	yes	yes	yes	yes

Notes: Panel A of the table reports coefficient estimates from instrumental variable regressions at the regional level in columns 4 through 6. Columns 1 through 3 report results from Ordinary Least Squares regressions. Standard errors are given in parentheses. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. In Panel A the dependent variable is the log of the ratio of mean per capita income in 1995 and 1989. The main variables of interest are the share of the region's 1961 population which are expellees who had (i) arrived from the Soviet sector or (ii) came directly to West Germany. We instrument for these variables with the share of the region's 1939 housing stock which was destroyed in 1946 and the amount of rubble per capita across the cities of a region in 1946. The first stage regressions corresponding to the instrumental variable regression in column 4-6 of Panel A are given in columns 4-6 of Panel B and C. All regressions control for a region's distance to the former East German border, the log of mean per capita income in 1989 and the log of the ratio of mean per capita income in 1989 and 1985. All regressions include 10 federal state fixed effects. Coefficient estimates for these controls are not shown for expositional clarity.

TABLE 5: SOCIAL TIES AND ENTREPRENEURIAL ACTIVITY

	(1)	(2)	(3)
	Income '95/'89 (p.c., log)		*Share*
	Entrepreneurs	*Non-Entrepreneurs*	*Entrepreneur 1995*
Share Expellees (Sov. S.) '61	4.008*	1.755**	0.322*
	(2.096)	(0.735)	(0.163)
Income Entrepreneurs '89 (p.c., log)	-0.657***		
	(0.194)		
Income Non-Entrepreneurs '89 (p.c., log)		-1.940***	
		(0.535)	
Share Entrepreneurs '89			0.496***
			(0.104)
R^2	0.578	0.662	0.794
N	71	71	71
Standard Controls	yes	yes	yes

Notes: The table reports coefficient estimates from instrumental variable regressions at the regional level. Standard errors are given in parentheses. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. The dependent variable in column 1 is the log of the ratio of mean per capita income of entrepreneurs in 1995 and 1989. The dependent variable in column 2 is the log of the ratio of mean per capita income of non-entrepreneurs in 1995 and 1989. The dependent variable in column 3 is the share of individuals who report in 1995 to be entrepreneur. The main variable of interest in all columns is the share of the region's 1961 population which are expellees who had arrived from the Soviet sector. We instrument for this with the share of the region's 1939 housing stock which was destroyed in 1946. First stage results are shown in Table 2. All regressions control for a region's distance to the former East German border, the log of mean per capita income in 1989 and the log of the ratio of mean per capita income in 1989 and 1985. All regressions include 10 federal state fixed effects. Coefficient estimates on these controls are not reported for expositional clarity. For details on the construction of the variables see data appendix. When column (1) and (2) are estimated simultaneously, we reject the equality of the coefficients on the share of expellees via the Soviet sector (p-value= 0.034).

TABLE 6: SOCIAL TIES AND FIRM INVESTMENT

	(1)	(2)	(3)	(4)	(5)
PANEL A: Reduced Form	*S. & B. in East Germany (Dummy)*				
Share Housing Destroyed '46	-0.030***	-0.029***	-0.029***	-0.028***	-0.031***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
S. & B. in West Germany (log)	0.119***	0.119***	0.119***	0.119***	0.119***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Distance to East (100km)		-0.013***	-0.013***	-0.012***	-0.014***
		(0.004)	(0.004)	(0.004)	(0.004)
Income 1989 (p.c., log)		0.001	0.004	0.017	0.001
		(0.032)	(0.033)	(0.031)	(0.032)
Income '89/'85 (p.c., log)			-0.016	-0.044	-0.008
			(0.043)	(0.048)	(0.044)
Sh. Employed in Manufact. '89				0.111	
				(0.084)	
Migration from East'91-'95					-1.517
					(1.125)
R^2	0.126	0.127	0.127	0.127	0.127
PANEL B: Second Stage					
Share Expellees (Sov. Sector) '61	1.579**	1.469**	1.556**	1.559**	1.616**
	(0.689)	(0.654)	(0.693)	(0.714)	(0.690)
PANEL C: Second Stage	*S. & B. in Poland (Dummy)*				
Share Expellees (Sov. Sector) '61	0.281**	0.290**	0.289**	0.298**	0.293**
	(0.137)	(0.133)	(0.140)	(0.145)	(0.137)
PANEL D: Placebo	*S. & B. in Old EU Countries (Dummy)*				
Share Expellees (Sov. Sector) '61	0.060	0.377	0.459	0.483	0.409
	(0.580)	(0.527)	(0.540)	(0.546)	(0.534)
PANEL E: Placebo	*S. & B. in New EU, exc. Poland (Dummy)*				
Share Expellees (Sov. Sector) '61	0.188	0.185	0.182	0.182	0.213
	(0.206)	(0.206)	(0.218)	(0.227)	(0.213)
PANEL F: Placebo	*S. & B. in Non-EU Countries (Dummy)*				
Share Expellees (Sov. Sector) '61	0.034	0.139	0.115	0.104	0.114
	(0.304)	(0.276)	(0.287)	(0.292)	(0.277)
N	19387	19387	19387	19387	19387

Notes: All regression report firm level regression results, using our sample of firms which are headquartered in West Germany and have at least one subsidiary or branch in West Germany. The standard errors are clustered at district level to account for likely spatial correlation. Panel A reports results from firm level Ordinary Least Squares regressions. Panel B-F report firm level instrumental variable regression results. The main variable of interest in these is the district level share of expellees via the Soviet sector. We instrument for this with the share of the district's 1939 housing stock which was destroyed in 1946. First stage results are not reported. The dependent variable in panel A and B is a dummy indicating whether a firm has a subsidiary or branch in East Germany. The dependent variables in panels C-F indicate whether a firm has a subsidiary or branch in the specified location. All regressions include 10 federal state fixed effects and 4 sector fixed effects. We control for distance to the former East German border at the district level. Log of per capita income in 1989, log of the ratio of per capita income in 1989 and 1985, the share working in the manufacturing sector 1989 and the sum of migrants from the East in 1991, 1993 and 1995 are regional level controls. All specifications in Panels B-E include the same controls as the respective specification in panel A. We do not report the results for expositional clarity.

TABLE 7: SECTOR SPECIFIC EFFECTS

	S. & B. in East Germany (Dummy)	N
Expellees '61 × Agriculture	3.382	313
	(4.310)	
Expellees '61 × Services	2.142**	15521
	(0.993)	
Expellees '61 × Manufacturing	0.156	3225
	(1.784)	
Expellees '61 × Government	-2.552	361
	(2.447)	
N	19420	
Instrument	Housing×Sector	

Notes: The table reports coefficient estimates and standard errors in parentheses from a firm level instrumental variables regression, using our sample of firms which are headquartered in West Germany and have at least one subsidiary or branch in West Germany. The standard errors are clustered at district level to account for likely spatial correlation. The main variable of interest is the interaction of 4 exhaustive sectorial dummies with the district level share of expellees via the Soviet sector. (The main effect of the share expellees via the Soviet sector is hence not included.) We instrument for these with the interaction of the sectorial dummies and the share of the district's 1939 housing stock which was destroyed in 1946. First stage results are shown in Table 2. The dependent variable is a dummy indicating whether a firm has a subsidiary or branch in East Germany. The regression includes 10 federal state fixed effects and 4 sector fixed effects. It also includes the same controls as the specifications in column 3 of Table 6. We do not report these results for expositional clarity. The second column shows the number of firms in each sector.

TABLE 8: HOUSEHOLD DATA

	(1) *All*	(2) *Ties*	(3) *No Ties*
Age '90	51.2	51.5	50.4
	(14.6)	(15.0)	(13.6)
Gender	0.29	0.33	0.22
	(0.46)	(0.47)	(0.41)
Years of Education '89	12.21	12.12	12.42
	(1.84)	(1.80)	(1.91)
Income 1989 (SOEP)	3304	3219	3492
	(1856)	(1935)	(1656)
Capital Income '89	783	799	746
	(1729)	(1867)	(1378)
Entrepreneur '89	0.046	0.045	0.047
	(0.209)	(0.207)	(0.212)
Not Employed '89	0.075	0.079	0.065
	(0.263)	(0.270)	(0.247)
N	1911	597	1314

Notes: Columns 1-3 show means and standard deviations in parentheses for our sample of households from the SOEP panel. We selected only households which were in the panel in all of 1985, 1989 and 1995. Income in 1989 and capital income in 1989 are reported in German Marks. The variables *Entrepreneur '89* and *Not Employed '89* are dummy variables indicating whether the household head is entrepreneur and not working, respectively. Column 1 shows data for all observations in our sample. Column 2 shows data for households with ties to relatives in East Germany. Column 3 shows data for households without ties to relatives in East Germany. Column 4 shows p-values of a t-test testing the equivalence of the means shown in column 2 and 3. See data appendix for details.

TABLE 9: REGION AND HOUSEHOLD LEVEL INCOME

	(1)	(2)	(3)	(4)	(5)
		Income '95/'89 (log)			
Level (Source)	*Aggregate (MZ)*	*Household (SOEP)*			
Share Expellees (Sov. Sector) '61	2.442***	2.777			
	(0.880)	(3.609)			
Ties to Relatives '91			0.069***	0.046**	0.049**
			(0.025)	(0.023)	(0.023)
Income 1989 (p.c., log, MC)	-0.209***	0.166	0.267**	0.260*	
	(0.060)	(0.186)	(0.126)	(0.137)	
Income '89/'85 (p.c., log, MC)	-0.355***	-0.649	-0.664	-0.798*	
	(0.086)	(0.446)	(0.449)	(0.450)	
Distance to East (100km)	0.011**	0.001	-0.009	-0.008	
	(0.004)	(0.025)	(0.016)	(0.017)	
Income 1989 (log, SOEP)		-0.242***	-0.248***	-0.340***	-0.338***
		(0.024)	(0.025)	(0.028)	(0.029)
Income '89/'85 (log, SOEP)		-0.115***	-0.117***	-0.144***	-0.146***
		(0.031)	(0.031)	(0.027)	(0.029)
Gender				-0.157***	-0.162***
				(0.024)	(0.024)
Age '90				-0.017***	-0.018***
				(0.005)	(0.005)
(Age '90)^2				0.000**	0.000**
				(0.000)	(0.000)
R^2	0.590	0.137	0.143	0.249	0.288
N	71	1911	1911	1911	1911
Fixed Effects	State	State	State	State	Region

Notes: Columns 1 shows results of a region level instrumental variable regression. Column 2-5 show results from household level regressions. In column 1 and 2 we instrument for the regional share of expellees via the Soviet sector with the share of the region's 1939 housing stock which was destroyed in 1946. First stage results are shown in Table 2. Columns 3-5 report results from Ordinary Least Squares regressions. The dependent variable in column 1 is the log of the ratio of per capita income in 1995 and 1989 at the regional level. The dependent variable in columns 2-5 is the log of the ratio of the household's income in 1995 and 1989. Columns 1 through 4 control for the same region level variables as column 3 in Table 3. Columns 2 through 5 control at the household level for the log of the household's income in 1989 and the log of the ratio of the household's income in 1989 and 1985. Columns 4 and 5 control additionally for the gender, age and age squared of the household head. The specifications in columns 1 through 4 include 10 federal state fixed effects. The standard errors are clustered on the regional level, which for columns 1 through 4 coincides with heteroscedasticity robust standard errors. Column 5 includes 71 region fixed effects. See data appendix for details on the construction of our variables.

TABLE 10: ROBUSTNESS

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Full Sample			*Income '95/'89 (log)*			
Ties to Relatives '91	0.049**	0.044**	0.047**	0.050**	0.049**	0.042*
	(0.023)	(0.021)	(0.023)	(0.023)	(0.023)	(0.021)
Years of Education '89		0.043				0.032
		(0.061)				(0.060)
(Years of Education '89)^2		0.000				0.001
		(0.002)				(0.002)
Capital Income '89 (log)			0.018***			0.014***
			(0.005)			(0.005)
Entrepreneur '89				0.058		0.025
				(0.071)		(0.065)
Not Employed '89					-0.028	-0.016
					(0.045)	(0.043)
R^2	0.288	0.326	0.294	0.288	0.288	0.331
N	1911	1911	1911	1911	1911	1911
PANEL B: Restricted Sample			*Income '95/'89 (log)*			
Ties to Relatives '91	0.055**	0.048*	0.054*	0.055**	0.055**	0.047*
	(0.027)	(0.025)	(0.027)	(0.027)	(0.027)	(0.025)
R^2	0.308	0.351	0.314	0.309	0.308	0.354
N	1490	1490	1490	1490	1490	1490
Household Level Controls	yes	yes	yes	yes	yes	yes

Notes: The table reports coefficient estimates from weighted least squares regressions at the household level. Standard errors, clustered at the region level to account for spatial correlation, are given in parentheses. The inverse of the sampling probability provided by SOEP is used as weights. Columns 1 replicates the results from the household level regression in column 5 in Table 9. The dependent variable is the log of the ratio of household income in 1995 and 1989. All specifications include the same controls as the specification in column 5 in Table 9, but we do not report results for expositional clarity. See data appendix for details on the construction of our variables. Panel A reports results using the full sample. In Panel B we replicate the regressions from Panel A using a restricted sample. In this sample we excluded households who did not have operational assets in 1989, but report to have such assets in 1995 as well as all households which report in either 1989 or 1995 to have income from renting out property. For expositional clarity we omit the results from co-variats.

TABLE 11: HETEROGENEOUS EFFECTS BY AGE OF HH HEAD

	Income '95/'89 (log)	N [with Ties]
Ties × Age Group below 40	0.092*	496 [153]
	(0.051)	
Ties × Age Group 40-51	-0.052	494 [162]
	(0.044)	
Ties × Age Group 52-62	0.108**	447 [151]
	(0.052)	
Ties × Age Group above 62	0.063*	474 [131]
	(0.037)	
R^2	0.448	
N	1911	

Notes: The table reports coefficient estimates from weighted least squares regressions at the household level. Standard errors, clustered at the region level to account for spatial correlation, are given in parentheses. The inverse of the sampling probability provided by SOEP is used as weights. The dependent variable is the log of the ratio of household income in 1995 and 1989. The explanatory variables of interest are the interactions of the dummy indicating ties to relatives in the East interacted with 4 exhaustive household head cohort dummies. The age quartiles are: 'below 40', 'between 40 and 51', 'between 52 and 62' and 'above 62'. The second column shows the number of household heads in each cohort. The number of household heads with ties to relatives in East Germany is given in square brackets. The specification also includes the same controls as column 5 of Table 9, so amongst others a full set of regional control. We do not report results for expositional clarity. See data appendix for details on the construction of our variables.

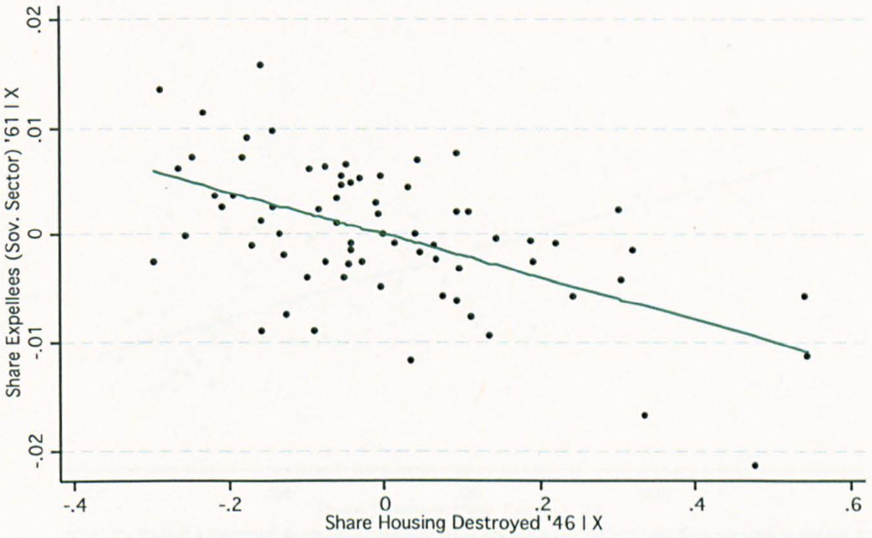
TABLE 12: HETEROGENEOUS EFFECTS BY INCOME

	(1)	(2)	(3)	(4)
		Income '95/'89 (log)		
Ties to Relatives '91	0.049**	0.046*	0.052**	0.050**
	(0.023)	(0.026)	(0.024)	(0.023)
Ties × Capital Income '89 (75th percentile)		0.011		
		(0.047)		
Capital Income '89 (75th percentile)		0.046		
		(0.028)		
Ties × Capital Income '89 (95th percentile)			-0.036	
			(0.112)	
Capital Income '89 (95th percentile)			0.067	
			(0.056)	
Ties × Capital Income '89 (99th percentile)				-0.022
				(0.242)
Capital Income '89 (99th percentile)				0.148
				(0.121)
R^2	0.288	0.290	0.288	0.289
N	1911	1911	1911	1911

Notes: The table reports coefficient estimates from weighted least squares regressions at the household level. Standard errors, clustered at the region level to account for spatial correlation, are given in parentheses. The inverse of the sampling probability provided by SOEP is used as weights. The dependent variable in all specifications is the log of the ratio of household income in 1995 and 1989. All specifications use ties to relatives in East Germany as explanatory variable and control for the same set of covariates as the specification in column 5 of Table 9, including a full set of region dummies. We do not report results for expositional clarity. Column 3, 4 and 5 control for a dummy indicating whether the household has capital income above the 75, 95 and 99 percentile of the capital income distribution, respectively. The explanatory variable of interest is the respective interaction of the capital income dummy and the dummy indicating ties to relatives in East Germany. See data appendix for details on the construction of our variables.

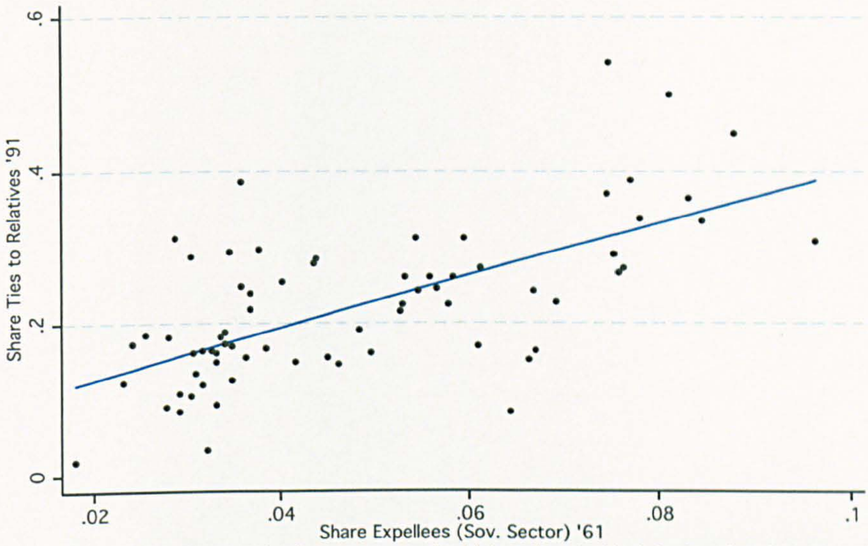
A.2 Figures

FIGURE 1: SHARE EXPELLEES AND SHARE HOUSING STOCK DESTROYED (CONDITIONAL SCATTERPLOT)



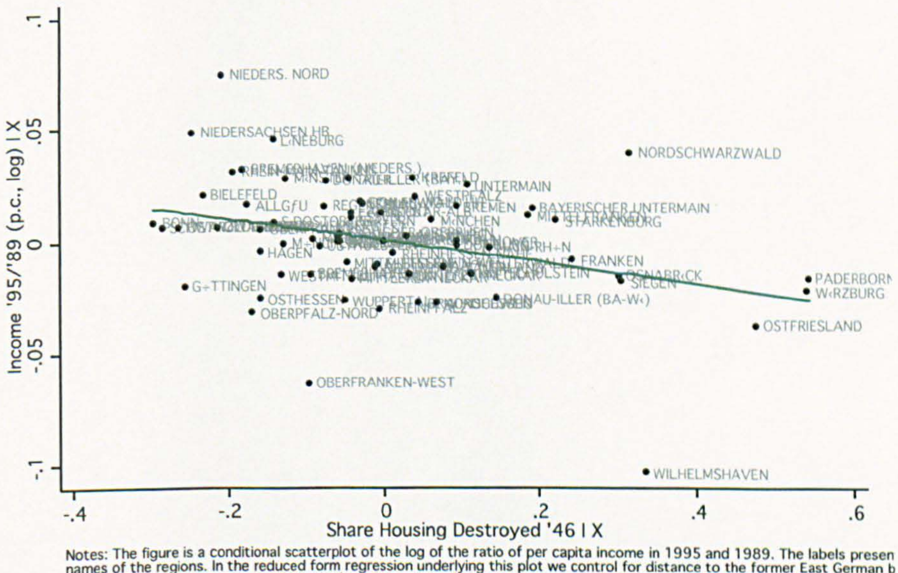
Notes: The figure is a conditional scatterplot of our measure of war destruction and the share of expellees at the regional leve In the first stage regression underlying this plot we control for distance to the former East German border, the log of per capi income in 1989, the log of the ratio of per capita income in 1989 and 1985 and a full set state fixed effects. Results from th regression are presented in column 2 of panel A of table 2. The solid line depicts the estimated linear relation between war de struction and the share of expellees.

FIGURE 2: SHARE TIES TO RELATIVES AND SHARE EXPELLEEES



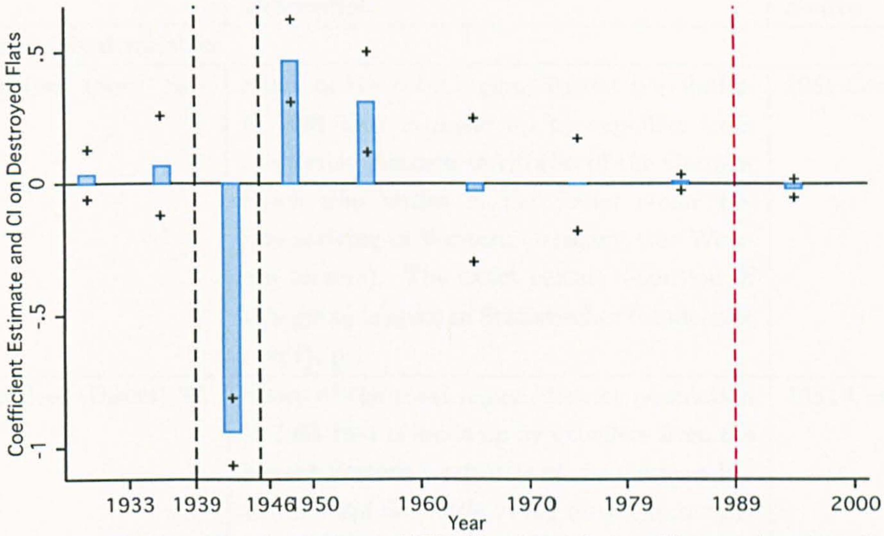
Notes: The figure is a scatterplot of our two measures of social ties at the regional level: the share of expellees and the share of individuals responding to have relatives in East Germany. The solid line depicts the estimated linear relations from an ordinary least squares regresssion of the share of expellees via the Soviet Sector on the share of individuals with relatives in the East and a constant. The coefficient estimate is 3.41 (robust s.e.=0.54) is significant at the 1% level.

FIGURE 3: INCOME GROWTH AND SHARE HOUSING DESTROYED (CONDITIONAL SCATTERPLOT)



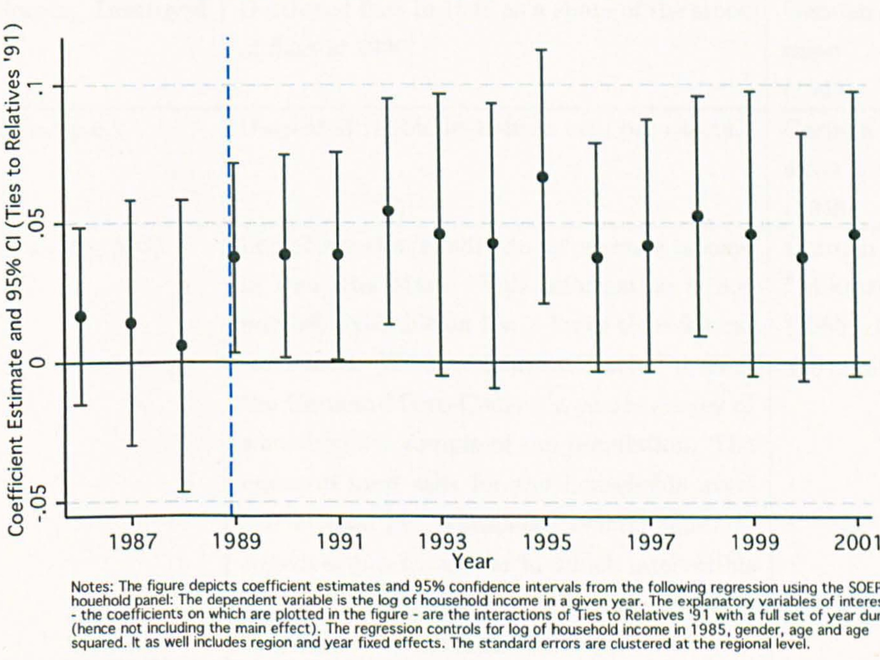
Notes: The figure is a conditional scatterplot of the log of the ratio of per capita income in 1995 and 1989. The labels presen names of the regions. In the reduced form regression underlying this plot we control for distance to the former East German b the log of per capita income in 1989, the log of the ratio of per capita income in 1989 and 1985 and a full set state fixed eff Results from this regression are presented in column 2 of panel C of table 2. The solid line depicts the estimated linear relatio between the log of the ratio of per capita income in 1995 and 1989 and war destruction.

FIGURE 4: EFFECT OF WWII DESTRUCTION ON POPULATION GROWTH



Notes: The figure depicts coefficient estimates and 90% confidence intervals for the coefficient on war destructions for 9 sep city level regressions. Each regression uses as dependent variable the population growth in between the dates specified on th horizontal axis (which are the years for which we have data) and includes as explanatory variables the share of expellees and a constant. The black dashed line indicate the period of WWII. The blue dashed line indicates the time of German reunification. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity.

FIGURE 5: INDIVIDUAL EFFECTS OVER TIME



Notes: The figure depicts coefficient estimates and 95% confidence intervals from the following regression using the SOEP houehold panel: The dependent variable is the log of household income in a given year. The explanatory variables of interest - the coefficients on which are plotted in the figure - are the interactions of Ties to Relatives '91 with a full set of year dumm (hence not including the main effect). The regression controls for log of household income in 1985, gender, age and age squared. It as well includes region and year fixed effects. The standard errors are clustered at the regional level.

A.3 Data

TABLE 13 - DATA DESCRIPTION AND SOURCES

Variable	*Description*	*Source*
PANEL A: Original variables		
Share Expellees (Sov. Sector) '61	Share of the total region/district population in 1961 that is made up by expellees from the former Eastern territories of the German Reich *who settled in the Soviet sector* before arriving in Western Germany (the Western sectors). The exact census definition of this group is given in Statistisches Bundesamt (1961), p.4.	1961 Census
Share Expellees (Direct) '61	Share of the total region/district population in 1961 that is made up by expellees from the former Eastern territories of the German Reich *who did not settle in the Soviet sector* before arriving in Western Germany (the Western sectors). The exact census definition of this group is given in Statistisches Bundesamt (1961), p.4.	1961 Census
Share Ties to Relatives '91	The respondents were ask whether they had relatives in the other part of Germany. We calculated the share of people who responded affirmatively.	SOEP (1991)
Subsidiaries and Branches in *loc*[a]	Number of subsidiaries and branches registered in *loc* belonging to the firm.	Orbis (2007)
Share of Housing Destroyed '46	Destroyed flats in 1946 as a share of the stock of flats in 1939.	German Association of Cities (1949)
Rubble '46 (m^3 p.c.)	Untreated rubble in 1946 in ccm per capita.	German Association of Cities (1949)
Income *t* (p.c., log, MC)	Log of a region's individuals' average income in Deutsche Mark. This information is not publicly available on levels lower than federal state level. We have hence extracted it from the German Micro-Census, a yearly survey of a random 1% sample of the population. The question used asks for the households average income per household member and respondent has to answer in which interval his household falls. We have taken the mean of the intervals as household income.	German "Mikrozensus" (1985, 1987, 1989, 1991, 1993, 1995)

Continued on next page

TABLE 13 – CONTINUED FROM PREVIOUS PAGE

Variable	*Description*	*Source*
Income Entrepreneurs t (p.c., log)	Log of region's individuals' average income in Deutsche Mark, for the subgroup of individuals who indicated to be 'entrepreneur' (with or without employees).	German "Mikrozensus" (1989, 1995)
Income Non-Entrepreneurs t (p.c., log)	Log of region's individuals' average income in Deutsche Mark, for the subgroup of individuals who indicated to have an occupation other than being 'entrepreneur' (with or without employees).	German "Mikrozensus" (1989, 1995)
Share Entrepreneur t	Regional share of individuals who indicate as occupation to be entrepreneur (with or without employees).	German "Mikrozensus" (1989)
Share Working in Manufacturing '89	Regional share of individuals who indicate to be working in manufacturing.	German "Mikrozensus" (1989)
Distance to East (100km)	Closest distance from a region's centre to the former GDR's border in 100 km.	- own calculations -
Migration from East '91-'95	Surveyed individuals who migrated to the region in the years 1991, 1993 and 1995 from East Germany as a share the total surveyed population.	German "Mikrozensus" (1991, 1993, 1995)
Region	The geocode of the 'spacial ordering unit', an agglomeration of on average 4.4 districts. This is the lowest level at which income data is made available in the German Micro Census.	
Ties to Relatives '91 (Dummy)	Dummy indicating whether household head or another person in the same household had relatives in the other part of Germany in 1991.	German SOEP (1991)
Income t (log, SOEP)	Log of income in German Mark of household head in year t .	German SOEP
Gender	Gender of highest ranked individual in the household for whom income data exists. Usually this will be the household head.	German SOEP (1990)
Age '90	Age of household head in 1990.	German SOEP (1990)
Years of Education '89	Years of education (including professional) of highest ranked individual in the household for whom income data exists. Usually this will be the household head.	German SOEP (1990)
Capital Income '89	Log of household capital income in 1989.	German SOEP (1990)

Continued on next page

TABLE 13 – CONTINUED FROM PREVIOUS PAGE

Variable	*Description*	*Source*
Occupation '89	We aggregated the occupations given in the German SOEP to the 8 categories 'Not Employed', 'Pensioner', 'In Education/Military Service', 'Worker', 'Farmer', 'White Collar', 'Entrepreneur' and 'Civil Servant'.	German SOEP (1990)
PANEL B: Generated variables		
Income t_1/t_0 (p.c., log)	Income t_1 (p.c., log) - Income t_0) (p.c., log).	
Share of Total Subsidiaries and Branches in *loc*[a]	Number of subsidiaries and branches in *loc* over the sum of this and the number of subsidiaries and branches in West Germany.	
Subsidiaries and Branches in *loc*[a] (Dummy)	Dummy variable that indicates whether the firm operates at least one subsidiary or branch in *loc*.	
Subsidiaries and Branches in *loc*[a] (log)	Logarithm of Subsidiaries and Branches in *loc*.	
Income t (p.c., log, SOEP)	Log of average income in the region, using the SOEP data.	
Income t_1/t_0 (log, SOEP)	Income t_1 (p.c., log, SOEP) - Income t_0) (p.c., log, SOEP)	

[a] Where *loc* stands for East Germany, West Germany, 'New' EU Countries, 'Old' EU Countries, and Non-EU Countries.

A.4 Further Results

TABLE 14: GMM USING PANEL STRUCTURE

	(1)	(2)	(3)
	Income (p.c., log)		
Share Expellees × 1995	2.813***	2.871***	2.538***
	(1.023)	(0.960)	(0.916)
Share Expellees × 1993	2.059**	2.117**	1.783*
	(1.022)	(0.959)	(0.915)
Share Expellees × 1991	1.532	1.590*	1.257
	(1.022)	(0.959)	(0.916)
Share Expellees × 1989	0.506	0.564	
	(1.022)	(0.961)	
Share Expellees × 1987	-0.329		
	(1.023)		
Income 1985 (p.c., log)	0.864***		
	(0.030)		
Income 1987 (p.c., log)		0.859***	
		(0.030)	
Income '87/'85 (p.c., log)		-0.487***	
		(0.059)	
Income 1989 (p.c., log)			0.871***
			(0.031)
Income '89/'85 (p.c., log)			-0.445***
			(0.055)
Distance to East (100km)	0.003	0.005**	0.004*
	(0.003)	(0.003)	(0.003)
N	563	492	421

Notes: The table reports coefficient estimates from an asymptotically efficient two-step GMM estimation. In the first step we applied the Bartlett kernel to estimate the covariances of the errors up to one lag. The dependent variable is in all regressions the log of mean per capita income in year *t*. The main variable of interest in all columns is the share of the region's 1961 population which are expellees who had arrived from the Soviet sector. We interacted this with a *full set* of possible year dummies (different across columns) and hence the main effect is not included. We instrumented for this with the interaction of the same year dummies with the share of the region's 1939 housing stock which was destroyed in 1946. All regressions control for a region's distance to the former East German border and include federal state-year fixed effects. Column 1 controls for the log of the mean per capita income in 1985, column 2 controls for the log of the mean per capita income in 1987 and column 3 controls for the log of the mean per capita income in 1989. Column 2 controls as well for log of the ratio of mean per capita income in 1987 and 1985. Column 2 controls for log of the ratio of mean per capita income in 1989 and 1985.

TABLE 15: SOVIET SECTOR AND DIRECT EXPELLEES

(A) SUMMARY STATISTICS (EXPELLEES, CENSUS '71)

	(1) *Ex. (Soviet Sector)*	(2) *Ex. (Direct)*	(3) *West Germans*
Years of Schooling '71	9.69	9.63	9.81
	(1.19)	(1.11)	(1.50)
Entrepreneur '71	0.03	0.03	0.06
Labour Force Participation '71	0.52	0.54	0.55
Primary Sector '71	0.04	0.05	0.12
Production and Construction '71	0.51	0.53	0.44
Services and Trade '71	0.33	0.30	0.32
Government '71	0.12	0.13	0.11
N	10120	49638	322240

(B) EXPELLEE SETTLEMENT AND REGIONAL CHARACTERISTICS '89

	(1)	(2)	(3)
	Coefficient		*p-value*
Outcome Variable	*Ex. (Soviet Sector)*	*Ex. (Direct)*	*(H_0: Equality of Coeff.)*
Years of Schooling '89	-0.398	-0.538	0.956
	(2.108)	(0.678)	
Share Entrepreneur '89	0.017	0.033	0.937
	(0.161)	(0.047)	
Share Unemployed '89	-0.130	-0.036	0.509
	(0.121)	(0.031)	
Sh. Employed in Agriculture '89	-0.406*	0.151*	0.059
	(0.240)	(0.081)	
Sh. Employed in Manufacturing '89	0.877	-0.022	0.350
	(0.806)	(0.222)	
Sh. Employed in Services '89	0.447	-0.195	0.338
	(0.558)	(0.162)	
Sh. Employed in Government '89	-0.323	0.005	0.395
	(0.355)	(0.066)	

Notes: Table 15 shows means, standard deviations in parentheses and the number of observations in square brackets. Data is from the 1971 edition of the German Census. Column 2 shows summary statistics for expellees via Soviet sector. Column 3 shows summary statistics for direct expellees. Column 1 shows data for all remaining individuals excluding refugees. Income in 1971 is given in German Marks. All other variables except years of schools are shares. The labour force participation and entrepreneurial share are given relative to the entire population. The sectorial distribution is given relative to all working individuals. Part B presents results from Ordinary Least Square regressions of the outcome variable shown in the leftmost column on the share of expellees (Soviet Sector), the share of expellees (Direct) and the same controls as column (3) of table 3. Each row represents an independent regression and we only report the coefficient estimates on the shares of expellees in column (1) and column (2). Standard errors are given in parentheses. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. Column (3) gives the p-value of a t-test of the equality of the coefficients in column (1) and (2).

TABLE 16: FIRM LEVEL DATA (SHARE)

	(1)	(2)	(3)	(4)	(5)
	Share of Total S. & B. in East Germany				
PANEL A: Reduced Form					
Share Housing Destroyed '46	-0.012**	-0.011**	-0.012**	-0.011**	-0.013***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
S. & B. in West Germany (log)	0.016***	0.016***	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Distance to East (100km)		-0.005***	-0.005***	-0.004**	-0.005***
		(0.002)	(0.002)	(0.002)	(0.002)
Income 1989 (p.c., log)		-0.008	-0.007	-0.003	-0.009
		(0.013)	(0.013)	(0.013)	(0.012)
Income '89/'85 (p.c., log)			-0.004	-0.014	0.000
			(0.019)	(0.020)	(0.018)
Sh. Employed in Manufact. '89				0.040	
				(0.035)	
Migration from East'91-'95					-0.805
					(0.492)
R^2	0.024	0.024	0.024	0.024	0.025
N	19387	19387	19387	19387	19387
PANEL B: IV					
Share Expellees (Sov. Sector) '61	0.646**	0.587**	0.619**	0.622**	0.657**
	(0.298)	(0.277)	(0.292)	(0.301)	(0.294)

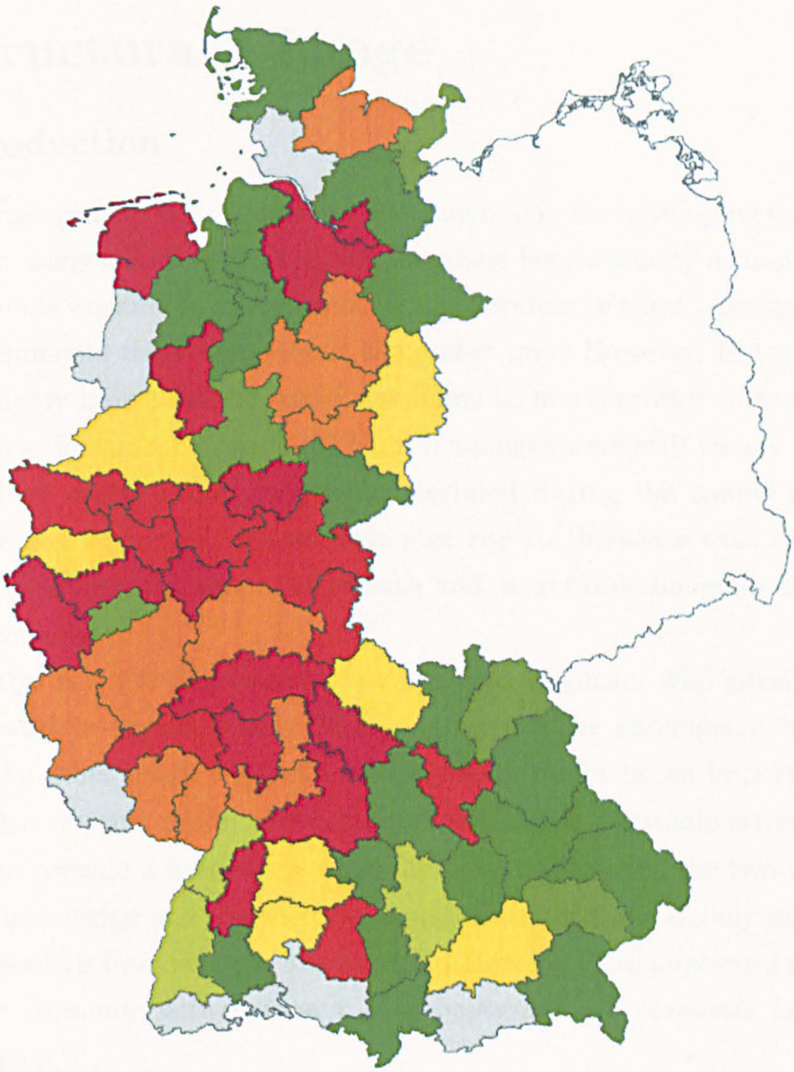
Notes: All regression report firm level regression results, using our sample of firms which are headquartered in West Germany and have at least one subsidiary or branch in West Germany. The standard errors are clustered at district level to account for likely spatial correlation. Panel A reports results from firm level Ordinary Least Squares regressions. Panel B reports firm level instrumental variable regression results. The variable of interest in these is the district level share of expellees via the Soviet sector. We instrument for this with the share of the district's 1939 housing stock which was destroyed in 1946. First stage results are not reported. The dependent variable is the share of a firm's subsidiaries and branches located in East Germany. All regressions include 10 federal state fixed effects and 4 sector fixed effects. We control for distance to the former East German border at the district level. Log of per capita income in 1989, log of the ratio of per capita income in 1989 and 1985, the share working in the manufacturing sector 1989 and the sum of migrants from the East in 1991, 1993 and 1995 are regional level controls. All specifications in Panels B include the same controls as the respective specification in panel A. We do not report the results for expositional clarity.

TABLE 17: EAST GERMANY

	(1)	(2)	(3)	(4)	(5)	(6)
			Income (log, SOEP)			
	1990	*1991*	*1992*	*1993*	*1994*	*1995*
Ties to Relatives '91	0.058	0.047	0.078*	0.046	0.068*	0.057
	(0.036)	(0.041)	(0.041)	(0.032)	(0.036)	(0.040)
Gender	-0.130***	-0.116***	-0.119***	-0.130***	-0.129***	-0.139***
	(0.024)	(0.028)	(0.025)	(0.026)	(0.028)	(0.030)
Age '90	0.067***	0.051***	0.038***	0.035***	0.026***	0.024***
	(0.005)	(0.005)	(0.007)	(0.006)	(0.004)	(0.005)
(Age '90)^2	-0.001***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.399	0.283	0.255	0.260	0.221	0.228
N	1506	1492	1473	1462	1474	1506

Notes: The table reports coefficient estimates from weighted least squares regressions at the household level. It uses the sample of households located in East Germany in both 1990 and 1995. The inverse of the sampling probability provided by SOEP is used as weights. Standard errors, clustered at the region level to account for spatial correlation, are given in parentheses. The dependent variable is the log of the household income in the specified year. The explanatory variable of interest is a dummy indicating ties to relatives in West Germany. All specifications include a full set of region fixed effects. See data appendix for details on the construction of our variables.

FIGURE 6: SHARE OF HOUSING DESTROYED



Notes: The figure presents the level of housing destroyed in 1946 in West German regions. The 5 colors refer to the 5 quintiles of war destruction, with red indicating those regions worst destroyed and green indicating the least destroyed regions. The cut-off values for the quintiles of housing destroyed are 0.093, 0.267, 0.377 and 0.526, respectively. The median level of housing destroyed in each quintile is 0.034, 0.189, 0.335, 0.406 and 0.591, respectively. Grey areas indicate regions for which we do not have data.

2 Language Heterogeneity and Structural Change

2.1 Introduction

Linguistic heterogeneity is a widespread phenomenon in the developing world. At the same time, in many countries with strong linguistic heterogeneity a small fraction of the population is working in manufacturing and services. Figure 7 presents a scatter plot that documents these facts about the status quo. However, historically much of the now highly industrialized world was home to much greater linguistic variety, too. In France, for example, some 20 local languages were still widely used at the end of the 19th century, but their usage declined during the course of the early 20th century (see Figure 8). These facts give rise to the sense that the evolution of language heterogeneity, economic growth and structural change in the economy might be connected.

In fact, this is a pre-dominant theme amongst linguists, who guestimate some 10% of the world's approximately 6000 languages to be endangered by economic forces. On the other hand, linguistic heterogeneity might be an important friction in the creation of joint productive activities and shape economic structures. This paper aims to provide a framework which helps to understand the two-way relation of language knowledge and the sectorial composition of the economy and economic growth. It uses this framework to answer two questions: *How important is a common language for economic activity?* and *How important are economic incentives for language change?*

In order to think about the first question, I outline and formalise how the stock of language knowledge might restrict the set of feasible joint economic activities. Key to this explanation is the idea that specialization of workers is at the heart of structural change. And specialized labor requires - in contrast to subsistence work - joint work and communication. In the absence of a common language economic agents cannot or only at a cost engage in joint economic activities and transactions - the stock of language knowledge hence shapes economic structures. The model presented in section 2.2 formalizes this notion. It is assumed that the ability to communicate is required for joint economic activities in a communication-intensive sector (referred to as 'manufacturing sector'), but no such requirement exists in a second sector (referred to as 'agricultural sector').[44] The formation of joint economic activities in the manufacturing sector is modeled by a stylized matching mechanism

[44]The idea that a common language allows for joint productive activities is obviously not new. It appears for example as early as in the story of the Tower of Babel in the Bible. Curiously enough, this story puts forward the idea that God created different languages with the only goal as to avoid such activites.

which allows for only a fraction of matches to be language-assortative. Speaking a common language hence plays the role of a matching friction in the formation of joint productive activities. The model clarifies what kind of 'language heterogeneity' we expect to be important and it yields testable predictions on the effect of the composition of language knowledge on both aggregate manufacturing growth and language group specific manufacturing growth.

Second, potential economic benefits from a wider pool of potential productive partners provide incentives to acquire language knowledge. Understanding this process is key to answering the second question. In section 2.3 I investigate the language learning decision. I propose a selection equation which depends on the potential economic benefits and on the heterogeneous (unobserved) net learning cost. Heterogeneity in the learning cost captures the differential abilities at learning a language and heterogeneity in the pleasure derived from this process. The benefit from language learning crucially depends not only on the own language group's size, but as well on the language learning decisions in all other language groups. I first discuss the simple case of two languages to gain some understanding of the basic mechanisms at work. I then present the case of many languages and provide a result for equilibrium existence. Further I show how a special, but interesting equilibrium can be found numerically and how the equilibrium depends on both the potential benefits of language learning and the share of activities which allow for language assortative-matching.

In order to think about the co-evolution of language knowledge and economic activities, section 2.4 endogenizes how the set of economic activities shape the incentives for language learning. In particular, it presents a competitive equilibrium model of consumption and production of two goods, produced by the two sectors. Amongst other things, it is an equilibrium outcome of the model that firms in the manufacturing sector have several employees, while in the agricultural sector firms are composed of a single worker only. This provides a micro-foundation for the assumption made in section 2.2, namely that one sector requires joint work, while the other does not. More importantly, this simple model clarifies how the sectorial composition determines the wage differential between the agricultural and manufacturing sector, and hence provides the missing link in the analysis of the joint determination of language knowledge and the sectorial composition of the economy. I then discuss the general equilibrium of language learning, the matching of workers into productive units, and the wage differential between the communication intensive and less communication intensive sectors.

The model highlights two key structural parameters, which can be interpreted as answering the two questions outlined above. In section 2.5 I provide tentative estimate these two parameters. I use data from a 1999 French survey of over 300.000 individuals, which allows to back out the language knowledge and change in French departements over the course of the 20th century. First, I estimate the share of activities which allow for language assortative matching. The parameter estimates suggest that a significant fraction of matches is not language-assortative, i.e. requires a common language for the feasibility of a joint economic activity. Secondly I estimate the mean language learning cost in units of the potential benefit of language learning, where the language to be learnt is chosen optimally. The estimates suggests that the mean language learning cost is around one eighths of the potential benefit of acquiring knowledge of an additional language.

To evaluate the general equilibrium model, I calibrate the remaining parameters of the model using its equations of state. I simulate the model and compare its predictions for the co-evolution of language knowledge and sectoral composition to the actual development from 1926 until 1946. I show that the model's predictions can explain at least 7% of the variation in the change of the agricultural share.

This paper contributes to the economic literature on the effects of fractionalisiation. These papers generally finds a negative association of measures of language fractionalisation and growth, investment or other economic variables (see Easterly and Levine (1997) for an early paper and Alesina and Ferrara (2005) for an excellent review paper). Much of this literature has interpreted these correlations as effects of *ethnic* fractionalisation on the functioning of markets, political institutions and conflict. This paper instead emphasises a possible direct effect of *language* diversity on economic growth.

The initial literature used data on ethno-linguistic groups from the Atlas Narodov Mira, which was compiled by Soviet researchers, to construct a Herfindahl-based index of ethno-linguistic fractionalization (ELF). Essentially this measures the probability of two randomly drawn individuals having their mother tongue in common.[45] Others proposed alternative ways to measure ethnic division, often referred to as 'polarisation' measures (see Esteban and Ray (1994) and their subsequent papers).[46] These have been shown to better explain the onset of ethnic conflict, but do not account for the effect of ethno-linguistic diversity on growth (Montalvo and Reynal-Querol, 2005b,a). This paper argues that this correlation might be driven by the simple fact, that *linguistic* diversity renders joint economic activities costly. The theory of this paper suggests how to measure the ease of communication. The measure proposed here explicitly accounts for bi- and multi-lingualism, which is key to understanding language change.

In doing so it also contributes to and draws on the economic and linguistic literature on the individual level economic returns to language proficiency. Notably, Barry Chiswick wrote a series of papers presenting results from various countries

[45] Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003) provided updated and redefined versions of this data.

[46] See Bossert, D'Ambrosio, and Ferrara (2010) for an alternative measure of fractionalisation.

(see e.g. Chiswick and Miller (1995)). These generally find strong positive returns to proficiency in the host countries main language amongst immigrants.[47]

This literature also puts forward a second empirical fact: the probability of being proficient in speaking and reading the host countries main language is negatively correlated with the size of the own language group (Chiswick and Miller, 1999). This paper contributes to this literature by providing a formal framework to think of the language learning decisions of individuals. I will point out in subsequent chapters how the model presented in this paper helps to organise our thinking about this literature. The closest relative of the model of language learning presented here is Lazear (1999). In fact, Lazear's model is a special case of the model I present. He interprets the model to analyse language learning of immigrant groups and an impressive list of related questions.

To the best of my knowledge this paper is the first analysis of the two-way relation of language learning and economic structures in a general equilibrium setting. An example of a linguists take on these issues is Coulmas (1992).

2.2 Languages and Manufacturing Growth

This section outlines a simple economic model to organize out thinking about (and estimate) the role of a common language for the formation of productive units. Key to this model is the assumption that if the labour force of a firm is greater than 1, workers need to be matched with a productive partner *with whom they are able to communicate*, i.e. with whom they have knowledge of a language in common.[48] I derive testable implications of the model.

2.2.1 The Environment

Sectors Workers live for T periods in region d. At the beginning of period 1 workers join one of two sectors in which they work for the remaining life time. Firms in these sectors produce goods f and m, respectively Assume free entry and exit of firms and competitive markets.

The two sectors are distinguished by their production technology, represented by a cost function. In both sectors production requires a single factor, labour. Firms in the first sector have a constant marginal labour cost, denoted by γ_d^f, and no fixed cost. Firms in the second sector have labour requirement for producing

[47]See, for example, Chiswick (1998) for evidence from Israel, Chiswick and Miller (1999) for evidence from the US, Dustmann and Fabbri (2003) for evidence from the UK and Chiswick, Patrinos, and Hurst (2000) or Godoy, Reyes-Garcia, Seyfried, Huanca, Leonard, McDade, Tanner, and Vadez (2007) for evidence from Bolivia.

[48]Note that this modelling assumption presumes that a common language is important for common work, but not for trading of goods. I belief this to be reasonable, as trading the goods might be done on an occasional basis.

x units of $\Gamma_d^m + \gamma_d^m x^2$ (i.e. Γ_d^m are the fixed and $\gamma_d^m x^2$ are the variable costs).[49] The parameters of the cost functions are heterogeneous across regions, indexed by d, capturing differences in the efficient production technology due to geographical, institutional or other factors.

With free entry, the equilibrium output in the second sector is $x^* = (\Gamma_d^m/\gamma_d^m)^{1/2}$ and a firm's equilibrium labour demand is $2\Gamma_d^m$.[50] The equilibrium size of a firm's labour force (measured in number of workers) is $2\Gamma_d^m$ over the labour supply of each worker in the second sector. This is generally different from 1. (I endogenize the labour supply in section 2.4.1 and derive an explicit formula for the equilibrium firm size in the second sector.) Conversely, in the first sector there is no benefit to producing jointly with co-workers. In summary, workers in the first sector, sector f, are self-employed, while workers in the second sector, sector m, work in productive units composed by several workers.

The model can be thought of as having one sector which generates increasing returns to scale by division of labour. This, in turn, requires groups of people to work together. We can think of the first sector as the 'agricultural' sector', producing 'food', and we can think of the second sector as the 'modern' or 'manufacturing and services' sector, producing 'manufacturing goods'. This is consistent with data from the 1926 French census, where 61.5% of all individuals in the agricultural sectors were head of the firm (farm), i.e. a majority of individuals was self-employed. In the industrial and services sectors the respective numbers are only 9.0% and 27.0%, respectively.[51]

Occupational Choice Denote with p_d^m and p_d^f the price of a unit of good m and f, respectively, in region d. Workers want to join the sector which offers the higher pay to each unit of labour supplied. The return to each unit of labour in the agricultural sector is p_d^f/γ_d^f. Denote with w_d^m the wage in the modern sector. For the moment make the following assumption (w_d and p_d^f will be endogenized in section 2.4):

Assumption 1 $w_d > p_d^f/\gamma_d^f$.

This implies that each worker wants to work in sector m.

Matching of Workers Crucially, to work in the modern sector, workers need to be matched with a co-worker *with whom they can communicate.* Workers can always choose to work in the agricultural sector where work is done individually. Assume the following matching process: Individuals joining the labour force at time t are matched randomly with another individual in region d joining the labour force.

[49] I denote endogenous variables with Latin characters and exogenous parameters with Greek characters.

[50] Free entry implies that each firm produces where average costs are minimised.

[51] Data from volume II/III, p. 164 of Statistique Générale de la France (1931b).

Assume each worker is matched with probability α with a worker from his own language group, and with probability $1 - \alpha$ with a worker from the region's full population. Hence $\alpha = 0$ is equivalent to a fully random, *non-assortative* matching process. Conversely, $\alpha = 1$ can be thought of as a perfectly *assortative* matching process. An estimate of α provides an answer to the question: How important are differences in language abilities for the formation of firms in the increasing returns to scale sector?

Language Groups The number of languages in region d is denoted by n_d and the indexed by $l = 1, 2, ..., n_d$. Any individual will speak one or several of these languages. At most there are $N_d := (2^{n_d} - 1)$ different combinations of language abilities. Let each such combination be called a 'language group', indexed by $j = 1, 2, \ldots, N_d$, and characterised by the $n_d \times 1$ vector $\boldsymbol{g}_{dj}$, with element l equal to 1 indicating knowledge of language l and 0 otherwise. The number of languages in common between any two individuals in language groups j and h is $\boldsymbol{g}_{dj}'\boldsymbol{g}_{dh}$. Hence those individuals are able to communicate if and only if $\boldsymbol{g}_{dj}'\boldsymbol{g}_{dh} > 0$. Let the fraction of language group j amongst the population of workers joining the workforce at time t be denoted by $s_{dj}(t)$. Further define $\boldsymbol{G}_d := [\boldsymbol{g}_{d1}, \boldsymbol{g}_{d2}, \ldots, \boldsymbol{g}_{dN_d}]$ and $\boldsymbol{s}_d(t) := [s_{d1}(t), s_{d2}(t), \ldots, s_{dN_d}(t)]'$.

2.2.2 Sectoral Growth

For a member of language group j the probability of having a language in common with a matched worker is

$$p_{dj}(t) := \Pr(\text{'successful match'}; \boldsymbol{G}_d, \boldsymbol{s}_d(t)) = \alpha + (1 - \alpha) \cdot \mathbf{1}(\boldsymbol{g}_{dj}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) \qquad (5)$$

where $\mathbf{1}$ is the indicator function on a vector, $\mathbf{1} : \mathbb{N}_0{}^{N_d} \rightarrow \{0, 1\}^{N_d}$, taking value 1 if the element of the vector is equal or greater than 1 (i.e. the two groups have at least one language in common) and 0 otherwise. Further, denote $\boldsymbol{p}_d(t) := [p_{d1}(t), p_{d2}(t), \ldots, p_{dN_d}(t)]'$. Then the fraction amongst the population joining the labour force in district d at time t which works in manufacturing is $\boldsymbol{s}_d(t)'\boldsymbol{p}_d(t)$. Conversely, the fraction of the population joining the labour force which works in agriculture, denoted $a_d(t)$, is

$$a_d(t) = 1 - \boldsymbol{s}_d(t)'\boldsymbol{p}_d(t) = (1 - \alpha) - (1 - \alpha) \cdot \boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d'\boldsymbol{G}_d)\boldsymbol{s}_d(t). \qquad (6)$$

This simple model provides three economically meaningful testable implications.[52]

[52] Obviously the fact that the agricultural share in the labour force alters, might effect both w_d and p_d^f. This will be modelled explicitly in section 2.4.1.

2.2.3 From Theory to Empirics

At the aggregate level, equation (6) predicts that the fraction of agricultural workers depends positively on language heterogeneity as measured by $\boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t)$ (*Testable Implication 1*). This can be tested with a regression of the form

$$a_d(t) = \gamma_0 + \gamma_1 \cdot \boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) + \epsilon_d^1. \tag{7}$$

The coefficient estimate $\hat{\gamma}_1$ measures the fraction of matches for which having a language in common is important. The regressor $\boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t)$ corresponds to the probability that any two individuals have a language in common.

Note that this can be understood as a generalised form of the Herfindahl-style index of mother-tongue linguistic fractionalisation which is frequently used in the literature on the economic effects of *'ethno-linguistic'* fractionalisation. It is generalised in the sense that it accounts for individuals who speak more than one language.[53] A simple example might serve to highlight the difference: Consider a region where some share z of the population speaks language 1 and 2, and the remaining population speaks language 2 only and suppose that for the first group language 1 is classified as 'mother tongue' (some unique language needs is always classified as mother tongue when indices of ethno-linguistic fractionalisation are constructed!) A Herfindahl-style index of 'ethno-linguistic' fractionalisation would generally be different for two regions with different z, while the measure of 'ease of communication' which the simple theory in this paper suggests would 1 independent of z, since all individuals can communicate with each other.

Equation (7) also makes precise how to estimate α and the conditions under which α is identified. In particular we require exogenous variation in $\boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t)$. Estimating (7) requires data on the fraction of agricultural workers *amongst those joining the labour force* at time t. Where this is unavailable we can estimate α with data on the aggregate share of workers in the agricultural sector at time t. Denote this by $A_d(t)$. The model implies $A_d(t) = \frac{1}{T}\sum_{s=t-(T-1)}^{t} a_d(s)$ or

$$A_d(t) = (1-\alpha)/T - (1-\alpha)/T \cdot \boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) + (1-1/T)A_d(t-1) \tag{8}$$

[53]The simple form of this – where each individual is assumed to have a single mother tongue – has been a prominent proxy for ethnic fractionalisation in the empirical development literature. This literature typically presents cross-country results from Ordinary Least Squares regressions of equation (7) using some measure of aggregate growth as outcome variable. The model presented here makes precise the reasons why this estimation strategy might be problematic: (i) the proxy of ethnic fractionalisation used is (at face value) really a measure of linguistic fractionalisation, and the interpretation of the correlations found might be quite different from the channels emphasised in this literature; (ii) linguistic fractionalisation might depend on the economic incentives to language learning which might be influenced by the sectorial composition in region d, and might hence be correlated with the error ϵ_d^1; and (iii) the simple measure of mother-tongue linguistic fractionalisation does not capture that individuals might speak several languages (or be members of two ethnic groups, for that matter) and might hence fail to explain some essential heterogeneity where bi-lingualism is prevalent.

where I assume $a_d(t-T) = A_d(t-1)$. Then α can be equivalently estimated from a regression of the form

$$A_d(t) = \delta_0 + \delta_1 \cdot \boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) + \delta_2 A_d(t-1) + \epsilon_d^2. \tag{9}$$

When T is unknown an estimate of α is given by $\hat{\alpha} = 1 + \hat{\delta}_1/(1-\hat{\delta}_2)$. Alternatively when T is known this can be estimated as constrained regression with $\delta_2 = 1 - 1/T$.

The values δ_0 and δ_1 in regression (9) satisfy $\delta_0 + \delta_1 = 0$ (*Testable Implication 2*). This hypothesis, if not rejected, implies that in a region where every worker can communicate with every other worker, i.e. $\boldsymbol{s}_d(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) = 1$, all workers would start working in the manufacturing sector. (Rejecting this hypothesis would shed doubts on whether Assumption 1 – that all individuals would want to work in the manufacturing sector, if they are matched with an individual which whom they can communicate – holds in the data.)

Thirdly, the model suggests some essential heterogeneity in the sectorial composition across language groups *within regions*. In particular, equation (5) implies that the probability of working in the manufacturing sector correlates negatively with the language group specific probability of having a language in common with a randomly matched worker, as measured by $\mathbf{1}(\boldsymbol{g}_{dj}{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t)$ (*Testable Implication 3*). This can be tested with a regression

$$a_{dj}(t) = \zeta_0 + \zeta_1 \cdot \mathbf{1}(\boldsymbol{g}_{dj}{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t) + \epsilon_{dj} \tag{10}$$

where $a_{dj}(t)$ is the share of agricultural workers amongst member of group j from region d joining the labour force at time t and ϵ_{dj} is a normally distributed error with variance inversely proportional to s_{dj}.[54]

When wages are higher in the manufacturing sector than in the agricultural sector[55], a simple corollary of this implication is that the average wage of members of a language group should increase with the group specific probability of being matched with a co-worker with whom he can communicate, $\mathbf{1}(\boldsymbol{g}_{dj}{}'\boldsymbol{G}_d)\boldsymbol{s}_d(t)$. One version of this is that immigrants who are proficient in their host country's main language should earn higher wages. Indeed, Chiswick and Miller (1999) report that earnings among legalized aliens in the US are higher by about 8% for men and 17% for women who are proficient in both speaking and reading English, compared to those lacking both skills. Chiswick (1998) find similar – if not stronger – correlations of income and proficiency in Hebrew and English amongst residents of Israel. And Godoy, Reyes-Garcia, Seyfried, Huanca, Leonard, McDade, Tanner, and Vadez (2007) report that fluency in spoken Spanish amongst the population of the Bolivian

[54] This is optimally estimated with the Generalised Least Squares estimator which takes the information about the error structure into account.

[55] This is assumed by Assumption 1.

Amazon was associated with 36.9-46.9% higher earnings compared with the earnings of monolingual speakers of the local language Tsimane'.[56] While these results are consistent with the implications of the theory, they provide at best an average treatment effect. They fail to exploit the potential heterogeneous treatment effects across immigrant groups suggested by the theory: Learning the host countries language is generally more beneficial for individuals of small immigrant groups that for members of very big immigrant groups. Neither do these papers report any evidence on the sectorial composition of labour across language groups.

I provide estimates of α and evidence on the testable implications in section 2.5.

2.3 Language Change

The previous discussion was concerned with the sectoral growth and its dependence on a *given stock of language knowledge*, characterised by $(\boldsymbol{G}_d, \boldsymbol{s}_d(t))$. This section discusses the evolution of language knowledge over time, i.e. the path of $\boldsymbol{s}_d(t)$. A complete model of language change needs to explain (i) how language knowledge is acquired and (ii) how language knowledge disappears. I focus attention on the effect of economic incentives on language learning. Section 2.3.1 discusses how individuals optimally choose acquiring a new language in response to economic incentives, given the set of language abilities inherited from their parents. Section 2.3.2 derives testable implications. This paper does not aim to discuss how languages are passed on across generations. This is, to some extend, driven by the belief that economic incentives are not of first-order importance for this decision.

2.3.1 Language Learning

Let each worker have the chance to learn one language before joining the labour force. The worker will do so if his expected benefit is greater than his cost of learning a language. Let this be some linearly additive utility cost c_{di}, which I take to be the realisation of a random variable with c.d.f. F_C, drawn independently and identically across individuals and regions. Assume $F_C(z) > 0$ for all $z > 0$ and $F_C(0) = 0$. Subscript i refers to the individual worker. The cost is measuring net forgone utility when learning a language. The assumed heterogeneity is meant to capture both differential talent at learning a language and heterogeneous pleasure derived from learning a language – in the end, many people find it a beautiful thing to do.

Denote with g_{dj}^l the vector characterising the language ability of a person endowed with language ability g_{dj} after learning additionally language l. (This corresponds to some g_{dh}.) Then the optimal language to learn for members of group j,

[56] Dustmann and Fabbri (2003) provide evidence suggesting that fluency in English increases the employment probability of black immigrants in the UK by about 22 percent points.

conditional on learning a language, is

$$l^*(j) := \arg\max_l \left[\mathbf{1}({g_{dj}^l}'G_d)\tilde{s}_d(t)\right] \tag{11}$$

where G_d and $\tilde{s}_d(t)$ characterise the language abilities of the population joining the labour force. Optimally, the individual learns the language which increases most the probability of being successfully matched. Denote with V_d^m and V_d^f the indirect utilities of working in the manufacturing and agricultural sector, respectively, and define $\pi := V_d^m - V_d^f$, i.e. π is the gain in indirect utility from working in the manufacturing sector over the agricultural sector. A worker decides to acquire language knowledge iff

$$(1-\alpha)\left[\mathbf{1}({g_{dj}^{l^*(j)}}'G_d)\tilde{s}_d(t) - \mathbf{1}({g_{dj}}'G_d)\tilde{s}_d(t)\right]\pi_d \geq c_{di}. \tag{12}$$

Learning a language has a positive net pay-off only if both $w_d > p_d^f/\gamma_d^f$ and learning a language increases the probability of being successfully matched sufficiently.[57]

The own language learning decision depends on the language learning of all other individuals of the same generation. Define an equilibrium as follows.

Definition 1 *A 'language learning equilibrium' is a vector $s_d(t)$ with elements $s_{dj}(t)$ s.t.*

$$\begin{aligned} s_{dj}(t) &= s_{dj}^0(t)\left[1 - F_C((1-\alpha)(\mathbf{1}({g_{dj}^{l^*(j)}}'G_d)s_d(t) - \mathbf{1}({g_{dj}}'G_d)s_d(t))\pi_d)\right] \\ &+ \sum_{\{h|g_{dh}^{l^*(h)}=g_{dj}\}} s_{dh}^0(t)\cdot F_C((1-\alpha)(\mathbf{1}({g_{dh}^{l^*(h)}}'G_d)s_d(t) - \mathbf{1}({g_{dh}}'G_d)s_d(t))\pi_d) \end{aligned} \tag{13}$$

Intuitively, an equilibrium is a vector $s_d(t)$ such that given this vector each individual takes his optimal language learning decision and these decisions result in aggregate in a language ability of the generation t characterised by $s_d(t)$. This is different from the $s_d^0(t)$ which characterised the language ability this generation had been endowed with by its parents.[58]

Generally there might be several language learning equilibria. I discuss results for two tractable cases: First, I present results for the case of two languages and F_C being the uniform distribution. In this case the equilibrium can be characterised in closed form. The discussion of this case will serve to gain some insights and

[57] Obviously language learning can have other than just wage pay-offs. For example it allows civic engagement, participation in social networks and access to education.

[58] Mathematically, this can be understood as follows: The first line is the share of the population endowed with language ability j times the fraction of this population which learns another language, and is hence no longer characterised by j. The second line is the sum over all those groups, which after taking their optimal language learning decisions have language ability j. The quantity being summed up is the fraction of each such group that learns the language which gives them language ability j, times the share of the population that this group made up before the language learning.

intuitions of the basic mechanisms at work. Second, I consider the case of many languages, but focus on a particular equilibrium: one in which all individuals who do not speak the language of the biggest language group learn this group's language and all members of the biggest language group learn the language of the second biggest language group. This equilibrium arguably characterises the French situation, it can be found numerically and I will focus on this equilibrium when calibrating the model.

Example 1: Two Languages Assume that only two languages exist in region d. Hence there are three language groups, $j \in \{1, 2, 3\}$ and $\boldsymbol{G}_d = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. The first group speaks only language 1, the second group speaks only language 2 and the third group speaks both languages. The decision *which* language to learn is trivial in this case. Any individual of language group 3 cannot learn a further language and any individual in group 1 or 2 would learn the language they do not speak. W.l.o.g. consider the language learning of an individual in group 1. He will learn language 2 iff

$$(1-\alpha)\left[1-(s_1+s_3)\right]\pi \geq c_i,$$

where we suppress the subscript d for notational simplicity. The term $1-(s_1+s_3)$ measures the differential probability of being able to communicate with a randomly matched worker: If the individual did learn the second language, he can communicate with workers from all language groups, whereas if he did not learn the other language, he can only communicate with individuals of language group 1 and 3. Obviously $1-(s_1+s_3) = s_2$, so in this simple case the benefit from learning language 2 is to be able to communicate with those who only speak this language, i.e. members of language group 2.

Three points are captured by this decision rule: i) the incentive to learn the other group's language decreases with the fraction of the other group who speaks ones own language; ii) individuals of the smaller language group will find it, ceteris paribus, more beneficial to invest in language learning; iii) with a high fraction of job matches being assortative on language knowledge, i.e. α is high, the incentive to learn a language is small.

An language learning equilibrium will be a vector $\boldsymbol{s} = \{s_1, s_2, s_3\}$ which solves simultaneously

$$s_j = s_j^0\left[1 - F_C((1-\alpha)\pi(1-(s_j+s_3)))\right]$$

for $j = 1, 2$. Since $s_3 = 1 - s_1 - s_2$ these are two equations in two unknowns. Assuming $c_i \sim U(0, \bar{c})$ allows to find a closed form solution. The following proposition characterises all equilibria.

Proposition 1 *Wlog assume $s_1^0 < s_2^0$. There are three types of equilibria:*

a. *Iff $s_2^0 < \frac{\bar{c}}{(1-\alpha)\pi}$ an interior equilibrium exists. It is unique and stable. The equilibrium value is given by*

$$\frac{s_j}{s_j^0} = \frac{\bar{c}^2 - (1-\alpha)\pi(\bar{c}/s_j^0)s_1^0 s_2^0}{\bar{c}^2 - (1-\alpha)^2\pi^2 s_1^0 s_2^0} \tag{14}$$

b. *Iff $s_1^0 < \frac{\bar{c}}{(1-\alpha)\pi} \leq s_2^0$ the unique equilibrium is $(s_1/s_1^0, s_2/s_2^0) = (0,1)$. It is stable.*

c. *Iff $\frac{\bar{c}}{(1-\alpha)\pi} < s_1^0$ then $(s_1/s_1^0, s_2/s_2^0) = (1,0)$ and $(s_1/s_1^0, s_2/s_2^0) = (0,1)$ are the only stable equilibria. (An unstable interior equilibrium exists.)*

The proposition highlights the role of the relative sizes of the groups, the potential benefit from learning a language and the possibility to match language assortatively. Let us first understand what different relative sizes of language groups imply for language learning.

If the two groups are similarly sized, cases a or c are likely to apply: either there is an interior equilibrium where a fraction of members of both groups learn the other group's language; or all members of one group learn the other group's language and nobody in the latter group does so, but both groups might be the one that learns the other group's language. In a loose sense, these are symmetric situations compared to case b, which is not suprising given that the groups are similarly sized.

In the equilibrium of case c all members of a group, say those who only speak language 1, learn the other group's language. This obviously removes any incentive for the latter group to learn the former group's language. And if, given that nobody who only speaks language 2 learns a language, everybody who only speaks language 1 has a sufficient incentive to learn language 2, this is an equilibrium. However, if $(1-\alpha)$ or the wage differential at stake is not too high, the interior equilibrium in a emerges and is stable.[59] In an interior equilibrium we have $s_1/s_1^0 < s_2/s_2^0$, i.e. in the smaller group a bigger fraction learns a second language.

If the groups are of unequal size, the only equilibrium is one where all members of the smaller group learn the language of the bigger group and no one in the bigger group learns the smaller group's language, corresponding to case b.

[59]Which of these types of equilibria emerges depends on the change in pay-off's when the other group's language learning changes. Say s_2 falls by a small amount. If the effect on s_1 is strong, then the share of bilinguals falls sharply, inducing s_2 to fall further. An interior equilibrium would be unstable and only a boundary equilibrium as in c would be stable. The equilibrium condition captures precisely under which conditions the response of s_1 to a small change in s_2 is big: if the importance of a common language is high (α is close to 0) or the wage differential at stake is high (π is big) the expected benefit of language skills for someone in who only speaks language 1 falls sharply as s_2 falls and hence s_1 rises strongly. The effect will as well be big if for the marginal worker the density of the distribution of language costs is high, so $\bar{c}$ is low.

The equilibria presented in Proposition 1 are partial equilibria, in the sense that they do not consider a potential effect of language learning on the wage differential. However, in situations where the effect of language knowledge on wages through the formation of manufacturing work relations is neglectable, this simple model might have predictive power or shed light on the empirical prevalence of bilingualism.

Example 2: Many Languages Now consider the case with $n_d > 2$ languages in region d (and omit subscript d again). Further assume that the initial language endowment characterised by $\boldsymbol{s}^0$ is such that everyone speaks just one language. The following proposition gives an existence result for an interesting equilibrium.

Proposition 2 *An equilibrium $\boldsymbol{s}^*$ exists such that all those member of language groups which are not the biggest language group learn, if they learn a language, the language of the biggest single language group. Any member of the biggest language group learns, if he learns a language, the language of the second biggest single language group. (Proof in Appendix B.1.)*

Proposition 2 captures the intuition that when many groups come together who initially do not have a language in common, it is an equilibrium for members of all groups other than the biggest to learn the same language. One can think of this as a 'dominant language equilibrium'. This is not necessarily the only equilibrium. W.l.o.g. let the order of $\boldsymbol{s}$ and $\boldsymbol{s}^0$ be such that the first n elements refer to the language groups which only speak one language and $s_1^0 > s_2^0 > ... > s_n^0$, element s_{n+1} refers to the language group speaking both language 1 and 2, and element s_{n+j-1}, $j = 3, 4, \ldots, n$, refers to the language group which speaks language 1 and j. Hence, by assumption, $s_j^0 = 0$ for all $j > n$. A necessary condition for an equilibrium as postulated in Proposition 2 is the existence of a vector $\boldsymbol{s}^*$ such that the following are satisfied:

$$s_1^* = s_1^0[1 - F_C((1-\alpha)\pi(s_2^*))] \tag{15}$$

$$s_2^* = s_2^0[1 - F_C((1-\alpha)\pi(s_1^* + \sum_{m=3}^{N_d}(s_m^0 - s_m^*))] \tag{16}$$

$$s_j^* = s_j^0[1 - F_C((1-\alpha)\pi(s_1^0 + \sum_{m\neq 1,j}(s_m^0 - s_m^*))] \, \forall j = 3, 4, \ldots, N_d. \tag{17}$$

From these it is easy to show the following corollary which describes the heterogeneity in language learning across differently sized groups:

Corollary 1 *In a language learning equilibrium as described in proposition 2 we have $s_1^*/s_1^0 > s_2^*/s_2^0 > ... > s_n^*/s_n^0$.*

This corollary states that the intensity of language learning is inversely related to the initial language group's size. In fact, there is evidence suggesting that indeed members of smaller immigrant groups learn the host countries language quicker and better. Chiswick and Miller (1999) report that the probability of being proficient in speaking and reading English is negatively correlated with the size of the own language group.[60] His coefficient estimates from a linear regression of a dummy of being proficient in English on the share of the own language group – a number in the interval $[0, 100]$ – range from -0.007 to -0.010 for males and from -0.006 to -0.013 for females. These are well in line with the theory presented here: A coefficient estimate of -0.010 would suggest that only when the own language group would make up the whole population, nobody would learn English.[61]

Further it is straight-forward to find the following corollary from the proof of Proposition 2.

Corollary 2 *Define a vector valued sequence $\boldsymbol{s}(n)$ where element $s_j(n)$, $j = 1, \ldots, n$, is the share of single language group j if individuals in j would make their optimal decisions taking $\boldsymbol{s}(n-1)$ as the language shares in the economy. Further take $\boldsymbol{s}(0) = [0, s_2^0, s_3^0, \ldots, s_N^0]$. Then $\boldsymbol{s}^*$ is the limit of $\boldsymbol{s}(n)$.*

Corollary 2 is interesting for two reasons: First, it shows that the equilibrium can be understood as the result of an dynamic, 'evolutionary' process. Second, it provides a converging, iterative method to find the equilibrium computationally. This will be useful when calibrating the model.

Lastly, we can show that the equilibrium as defined in Proposition 2 and computed in Corollary 2, satisfies two important comparative static properties:

Proposition 3 *For any $\pi > \pi'$ the equilibrium computed in Corollary 2 satisfies $\boldsymbol{s}^*(\pi, \alpha) \leq \boldsymbol{s}^*(\pi', \alpha)$. Similarly, for any $\alpha > \alpha'$ we have $\boldsymbol{s}^*(\pi, \alpha) \geq \boldsymbol{s}^*(\pi, \alpha')$. If at least one element $s_j^*(\pi, \alpha) \in (0, s_j^0)$, then these hold with strict inequality for at least one element, respectively. (Proof in Appendix B.1.)*

This result shows that with an increase in the potential benefit of working in the manufacturing sector (or an decrease in the fraction of the population who can match assortatively), a bigger fraction of each individual group would learn a language. In the case of an increaseed benefit of working in the manufacturing sector, this will imply that the probability of a successful match increases. This is not true for a change in α: while the amount of language learning increases, this might be offset by the higher probability of mismatches due to less assortative matching.

[60] As well Chiswick (1998) reports evidence from Israel that "Hebrew fluency increases with [...] living in an area in which a smaller proportion speak one's mother tongue."

[61] Obviously these can hardly be interpreted as causal as various selection mechanisms might be at work. The point is that these facts are consistent with the model's implications.

2.3.2 From Theory to Empirics

The model of language learning has several testable implications, and I have touched upon some of them in the previous section. Other implications are, for example, that equation (11) predicts which language will be learnt, conditional on a language being learnt. This prediction could be tested with a multinomial logit and one could compare the models predictive power against likely alternative models. Further, equation (12) gives rise to a selection equation, which potentially can be tested.[62]

This paper will focus on an implication of the model which allows to estimate a parametrised version of the distribution F_C. This (a) allows to understand what language learning policies might be able to achieve and (b) provides an important structural parameter to calibrate the general equilibrium model in section 2.6. To this end, make the following assumption:

Assumption 2 *Assume the distribution of language learning costs is exponential, i.e.*

$$F_C(c;\lambda) := \Pr(c_{di} < c;\lambda) = 1 - e^{-\lambda c}.$$

This distribution satisfies the assumption of positive support and $F_C(c) < 1$ for all $c > 0$. The parameter to be estimated, λ, is the inverse of the mean language learning cost. Under Assumption 2, equation (15) can be rewritten as

$$\log s^*_{d1} = \log s^0_{d1} - \lambda(1-\alpha)\pi_d s^*_{d2}. \tag{18}$$

This suggests that λ can be estimated within the linear regression framework. Both π_d and s^*_2 are endogenous. The system of simultaneous equations defined by (15)-(17) suggests that s^0_{d2} is a suitable excluded instrument for $\pi_d s^*_{d2}$. Taken at face value, I can estimate λ from a regression of the form

$$\log s^*_{d1} = \theta_0 \log s^0_{d1} + \theta_1 \widehat{\pi_d s^*_{d2}} + \theta_1(\pi_d s^*_{d2} - \widehat{\pi_d s^*_{d2}}) + \epsilon^3_d. \tag{19}$$

where $\widehat{\pi_d s^*_{d2}}$ is the predicted value from the first stage. One can back out the estimate of λ as $\lambda = -\theta_1/(1-\hat{\alpha})$ where $\hat{\alpha}$ is the estimate obtained in section 2.2.3. The model predicts that $\theta_0 = 1$ and $\theta_1 < 0$ (*Testable Implications 4 and 5*).

2.4 General Equilibrium

2.4.1 A Simple Economy

I now specify a simple two-sector model which pins down how different shares of manufacturing and agricultural workers map into wages and prices and hence eco-

[62] If an instrument for the cost of language learning is available, this equation can be used to estimate the marginal treatment effects as defined by Heckman and Vytlacil (2005) and for example calculate the treatment effects of different language policies.

nomic incentives to learn a language. Let individual preferences be represented by the utility function

$$U = \beta \log f_{di} + m_{di} - (\eta/2)e_{di}^2 \tag{20}$$

where f_{di} is the amount of food consumed, m_{di} is the amount of manufacturing goods consumed and e_{di} is the labour time supplied. The first bit are standard log-linear preferences on consumption and the second bit is a convex effort cost of providing labour. β and η are preference parameters.

I present the assumptions about the firms' production technologies in section 2.2. Assume free entry and exit, competitive markets and costless trade of manufacturing goods across and within regions, but agricultural goods can only be trades within regions.

Choices

Marshallian Demands: Given income y_{di} the Marshallian demands for m and f are

$$(f_{di}^*, m_{di}^*) = \begin{cases} (\frac{y_{di}}{p_d^f}, 0) & \text{if } y_{di} < \beta p_d^m \\ (\frac{p_d^m}{p_d^f}\beta, \frac{y_{di}}{p_d^m} - \beta) & \text{otherwise.} \end{cases} \tag{21}$$

The utility function hence represents preferences such that all income (in units of the manufacturing good) up to some critical income level is spend on food. Income above this level is spend on manufacturing goods.

Labour Choice and Income: Denote with w_d^s the income from one unit of labour in sector s.[63] Using the Marshallian demands we find the indirect utility $V(y_{di}, e_{di})$, given labour supply and income. The optimal labour supply, e_{di}^*, is defined by the first order condition (applying the Envelope Theorem)

$$V_1(y_{di}, e_{di}^*)w_d^s - \eta e_{di}^* = 0 \tag{22}$$

where the subindex 1 indicates the partial derivative with respect to the first argument.

Food Supply: The supply of agricultural goods is perfectly elastic where $p_d^f = w_d^f \gamma_d^f$. The amount a farmer produces depends on the marginal utility of income. The indirect utility of income given e_{di} is

$$V(y_{di}, e_{di}) = \begin{cases} \beta \log\left(\frac{y_{di}}{p_d^f}\right) - (\eta/2)e_{di}^2 & \text{if } y_{di} \leq \beta p_d^m \\ \beta \log\left(\frac{p_d^m}{p_d^f}\beta\right) + \frac{y_{di}}{p_d^m} - \beta - (\eta/2)e_{di}^2 & \text{otherwise.} \end{cases} \tag{23}$$

[63] The un-superscripted w is hence the same as w^m.

For a self-employed farmer the wage income is the same as the value of his produce, i.e. $y_{di} = p_d^f e_{di}/\gamma_d^f$. Given the indirect utility of a farmer in (23), there are two cases for the optimal individual labour choice e_{di} as defined by (22).

Case 1: Suppose that $y_{di} \leq \beta p_d^m$. Then we have $e_{di}^* = (\beta/\eta)^{\frac{1}{2}}$. This implies $y_d(e_{di}^*) = (p_d^f/\gamma_d^f)(\beta/\eta)^{\frac{1}{2}}$. For consistency we need that $y_d(e_{di}^*) \leq \beta p_d^m$ or $(p_d^f/p_d^m) \leq (\beta\eta)^{\frac{1}{2}}\gamma_d^f$. This is the 'subsistence case', i.e. where farmers consume what they produce. There is no supply of food in this case.

Case 2: Suppose that $y_{di} > \beta p_d^m$. Then $e_{di}^* = p_d^f/(p_d^m \gamma_d^f \eta)$ and $y_d(e_{di}^*) = (p_d^f/\gamma_d^f)p_d^f/(p_d^m\gamma_d^f\eta)$. For consistency we need $(p_d^f/p_d^m) > (\beta\eta)^{\frac{1}{2}}\gamma_d^f$. In this case the excess supply of food by each farmer is $E_{di}(f) = p_d^f/(p_d^m(\gamma_d^f)^2\eta) - \beta p_d^m/p_d^f$.
The conditions for the cases to be internally consistent are exhaustive and mutually exclusive. Note however, that this presumes the existence of a well-defined price-vector, which is not guaranteed in the subsistence case as no food is supplied to the market! This point will be further discussed below.

Manufacturing Supply: Firms which produce manufacturing goods supply the quantity where price is equal to marginal cost as long as the price is bigger than minimal average costs.

Equilibrium

Equilibrium Prices: Given A_d (suppressing the time dependence for the moment), an equilibrium in the goods market is a price vector such that excess supply of food by farmers equals the demand for food by non-farmers (the second market will clear by Walras Law), or:

$$\left(\frac{p_d^m}{p_d^f}\right)^* = \sqrt{\frac{A_d}{(\gamma_d^f)^2\eta\beta}}. \tag{24}$$

If A_d increases, total food production increases. For the markets to clear, p_d^f/p_d^m has to fall, decreasing the production and increasing the demand of food. Note that for every $A_d < 1$ the condition for a non-subsistence equilibrium is always satisfied, i.e. given that some of the labour force works in manufacturing, the equilibrium price vector will be such that agricultural workers supply food to the market, i.e. there is an 'exchange equilibrium'.[64]

[64] However, this is only true when some of the labour force works in the manufacturing sector. If not, i.e. $A_d = 1$, the condition for the subsistence equilibrium is satisfied. This makes trivial sense, as all workers produce food with the same technology and consume what they produce as there are no gains from trade. Note that the interpretation of this when A_d is endogenous less trivial! Suppose a subsistence equilibrium exists. How does the transition to a market exchange equilibrium happen if individuals are price takers, but prices for 'new goods' do not yet exist? The model (and I imagine any model with perfect competition) has nothing to say on this. The answer

Equilibrium Wages: With free entry firms open/exit in the manufacturing sector, equilibrium wages are such that each firm produces where average costs are minimised, i.e. $m_d^* = (\Gamma_d^m/\gamma_d^m)^{1/2}$. The zero-profit condition (all income is labour income) pins down the equilibrium ratio of wages and the price of manufacturing goods: $p_d^m m_d^* - w_d^m \left[\Gamma_d^m + \gamma_d^m m_d^{*2}\right] = 0$. As $m_d^* = (\Gamma_d^m/\gamma_d^m)^{1/2}$ we have

$$\left(\frac{p_d^m}{w_d^m}\right)^* = 2\left(\Gamma_d^m \gamma_d^m\right)^{1/2}. \tag{25}$$

Equilibrium Firm Size The equilibrium labour demand of a firm is $\Gamma_d^m + \gamma_d^m (m_d^*)^2 = 2\Gamma_d^m$. We know that the labour supply of a worker in the manufacturing sector is $e_d^* = w_d^m/(p_d^m \eta)$. Hence the equilibrium number of workers in a firm, firm size FS_d, will be

$$FS_d = 2\Gamma_d^m \left(\frac{p_d^m}{w_d^m}\right)^* \eta. \tag{26}$$

Think of technological progress as a change in the production function, which implies an optimal output expansion path such that Γ_d^m increases and γ_d^m decreases. This has two effects: There is a direct effect of Γ_d^m which increases labour demand and an indirect, ambiguous effect of Γ_d^m and γ_d^m on p_d^m/w_d^m which changes labour supply.

Language Learning Incentives: We are now equipped to determine how the share of manufacturing workers, A_d, maps into the incentives to acquire new language knowledge. The difference in utilities from working in the manufacturing and the agricultural sector is [65]

$$V_d^m - V_d^f = \frac{1}{2\eta (p_d^m)^2}\left((w_d^m)^2 - (p_d^f/\gamma_d^f)^2\right). \tag{27}$$

Evidently, this is positive iff $w_m > p_f/\gamma_f$, or if the pay-off from one unit of labour for manufacturing work is higher than the pay-off from one unit of agricultural work.[66] Using the previous results we can express $V_d^m - V_d^f$ when markets are in equilibrium in terms of A_d and exogenous parameters only:

$$V_d^m - V_d^f = (8\gamma_d^m \Gamma_d^m \eta)^{-1} - (\beta/2)A_d^{-1}. \tag{28}$$

Hence $V_d^m - V_d^f$ is increasing in A_d.

---

would presumably require to think of entrepreneurs and how they form expectations about prices in a not yet existing equilibrium.

[65]The income of a manufacturing worker is $w_d^m e_m^*$. The indirect utility of a self-employed agricultural worker (with $w_d^f = p_d^f/\gamma_d^f$) is calculated similarly.

[66]Note that the difference in indirect utilities is a measure of the income difference *in units of manufacturing goods.*

2.4.2 The General Equilibrium

A Benchmark As a benchmark, let us briefly discuss the equilibrium in this economy if no matching frictions existed in the formation of manufacturing units. Then workers flow in and out of the agricultural sector freely. At an interior equilibrium it needs to hold that $V_d^m - V_d^f = 0$. This is satisfied at $\overline{A}_d := 4\gamma_m \Gamma_m \eta\beta$. For any $A_d < \overline{A}_d$ the pay-off from agricultural work is higher, and for $A_d > \overline{A}_d$ the pay-off from manufacturing work is higher, such that this is a stable equilibrium. Further $\overline{A}_d > 0$, i.e. there will always be an agricultural sector. For a manufacturing sector to exist we need that $\overline{A}_d = 4\gamma_d^m \Gamma_d^m \eta\beta < 1$. Hence manufacturing costs need to be low enough for a non-subsistence equilibrium to exists.[67] [68]

General Equilibrium Consider a region which is in a language learning equilibrium as characterised by Proposition 2. At time t the generation which retires has agricultural share $a_d(t - T)$. Define a general equilibrium of language learning and manufacturing growth as:

Definition 2 (General Equilibrium) *A General Equilibrium in region d is a vector $\boldsymbol{s}_d^*$, an agricultural share A_d, and a utility differential π such that equations (8), (15)-(17) and (28) are satisfied simultaneously, given $\boldsymbol{s}_d^0$ and $A_d(t-1)$.*

We find that for the generation which joins the labour force there is a unique equilibrium of manufacturing growth and language learning.

Proposition 4 *There is a unique general equilibrium equilibrium. With $\overline{A}_d$ low enough, individuals strictly prefer working in the manufacturing sector conditional on being matched and a fraction of the population becomes bilingual. The fraction of the population which becomes bilingual is increasing in $A_d(t-1) - a_d(t-T)/T$ and λ. It is decreasing in β and $\gamma_d^m \Gamma_d^m \eta$.*

The intuition of these results is simple: When $\overline{A}_d$ is lower than the current value of A_d, then individuals want to work in the manufacturing sector, since it has a higher pay-off. The reason that A_d is relatively high is that individuals cannot work in the manufacturing sector when not successfully matched with a co-worker with whom they have a language in common. The incentives for language learning are higher if fewer people are already working in manufacturing, i.e. the higher is the fraction of the population which remains in the workforce, that works in the

[67]Incidentally, the condition implies that technological progress of the form that marginal costs decrease at the same proportional rate as fixed costs increase is not enough to 'escape' the subsistence equilibrium.

[68]Obviously, food cannot weigh to highly in the utility function and the marginal cost of providing labour, η, should not be too high as this makes working in manufacturing unattractive. The same is true for working in agriculture obviously, but there increases in η are compensated for by a more favourable goods price ratio.

agricultural sector; if the manufacturing sector is more productive, i.e. $\gamma_d^m\Gamma_d^m$ is lower; and if the relative appreciation of manufacturing goods is lower. A lower A_d directly translates into a higher GDP, as long as wages in the manufacturing sector are higher than in the agricultural sector.

The fact that incentives for language learning are higher if fewer people are working in manufacturing has an interesting implications for the dynamics of language learning over time. Consider a situation where in each new generation a fraction $\overline{A}_d$ of the population becomes agricultural workers, there are no incentives for language learning, and the remaining population can work in manufacturing without any language learning being undertaken. Now imagine a technological innovation such that $\overline{A}_d$ decreases. Then the incentives for language learning are highest in the first generation to join the labour force and decrease thereafter. Hence, we would expect technology driven language change to be strong initially, but with decreasing intensity later on. This is driven by the price mechanism, which lets the relative price of agricultural goods increase as fewer individuals work in agriculture.

2.5 Estimation

This section provides estimates of the structural parameters α and λ and tests the model's implications 1-5. Both parameters have a well-defined economic meaning.

2.5.1 Data

I use data from the survey 'Etude de l'Histoire Familiale', conducted in 1999 by the French statistical institute INSEE. This survey covers over 370.000 individuals and is linked to the French Census data. Amongst other things, the survey enquires about the history of an individual's language use with the following questions: 'Which language did/do you mainly speak with your r?' where r are all of 'mother', 'father', 'children', and 'partner'. Similarly, it asked 'Which language did/do you occasionally speak with your r?' Individuals could give several answers to each question. I use data from individuals born between 1906 and 1926, roughly those individuals who join the labor force between 1926 and 1946.

I use the data to construct the indicator vector $\boldsymbol{g}_{di}$ characterizing which language an individual i in départment d is able to speak. The nth element is 1 if individual i names language n in response to any of the above questions and 0 otherwise. Individuals with identical $\boldsymbol{g}_{di}$ form 'language group' j and s_{dj}^* is the share of the population that speaks only the language(s) of group j. This allows to straightforwardly calculate the probability that any two randomly drawn individuals in a départment d have a language in common, $\boldsymbol{s}_d^*(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d^*(t)$, and the probability that a member of language group j in district d is matched with someone he can communicate with, $\mathbf{1}(\boldsymbol{g}_{dj}{}'\boldsymbol{G}_d)\boldsymbol{s}_d^*(t)$. I refer to those as *Aggregate Matching Probability* and *Group Matching Probability.*

I construct from the data two other measures of language heterogeneity, both of which do not account for languages acquired after the childhood. The first repeats the calculation of the *Aggregate Matching Probability* but only considers language spoken 'mainly' with the parents. This is a measure of language knowledge and heterogeneity if no language learning took place subsequent to the childhood. I refer to it as *Parent Tongue Matching Probability* and take it to be the empirical counterpart of $\boldsymbol{s}_d^0(t)'\mathbf{1}(\boldsymbol{G}_d{}'\boldsymbol{G}_d)\boldsymbol{s}_d^0(t)$. The second measure repeats the calculation of the *Aggregate Matching Probability* but only considers the first language spoken 'mainly' with the mother (or father if no data on the mother is available), referred to as *Single Mother Tongue Matching Probability.* This measure is the analogue of the commonly used Hirschman-Herfindahl-Index of 'ethno-linguistic fractionalization'.

Further I construct two measures which are closer related to what people understand as ethnic heterogeneity: I obtained data on the shares of foreign nationals, for each nationality, in the population of a departement in 1926. I calculate from this the simple share of foreign nationals, *Share Foreigners,* as well as the Hirschman-Herfindahl-Index for nationalities (including French), referred to as *Hirschman-Herfindahl Nationalities.*

Lastly I use data on the share of the active working population with an occupation in agriculture, forestry or fishing in 1926 and 1946 from the French Census (*Share Agriculture t*). I believe these correspond most closely to what characterizes the 'agricultural' sector in the model, namely that little communication with co-workers is required during the work. As mentioned previously, this is also consistent with data from the 1926 French census, where 61.5% of all individuals in the agricultural sectors were head of the firm (farm), i.e. a majority of individuals was self-employed while in the industrial and services sectors only 9.0% and 27.0%, respectively, had this role.[69] I obtained data on agricultural and manufacturing wages in 1929, as well as the distribution of firm sizes. The regional data covers 78 French départements in 22 regions.

2.5.2 Language Heterogeneity and Sectorial Composition

In order to estimate α, Table 19 presents Ordinary Least Squares and instrumental variable regression results for the structural equation given by (9). The dependent variable is Share Agriculture 1946. Column 1 of Table 19 presents coefficient estimates and standard errors from running a simple Ordinary Least Squares regression. The coefficient estimate on the conditional matching probability is -0.108 (s.e.= 0.049), negative and significant. Taken at face value the coefficient estimate implies $\alpha = 0.079$. This suggests that some 8% of productive matches are language-

[69]Data from p. 164, volume II/III of Statistique Générale de la France (1931b).

assortative, i.e. individuals can search easily for a productive partner in their own language group. However, this calculation uses $\hat{\delta}_2 = 0.882$, which implies that $T = 8.3$. Given that *one* time period is 20 years, this clearly is unreasonably high. This points to the existence of some important factor shaping the persistence of the département specific agricultural share, other than language heterogeneity. I assume for further calculations therefore $T = 2$, which implies 40 working years after joining the labour force. Obviously the fact that $\hat{\delta}_2$ is biased sheds doubts on whether our estimate of δ_1 is as well unbiased. As a first check I constrain the coefficient estimate on the agricultural share to be 0.5, as implied by $T = 2$. Results from a constraint regression yield a coefficient estimate for δ_1 of -0.088 (s.e.=0.113), hence it is in a very similar ball-park as the unconstrained estimate. Assuming $T = 2$ implies an estimate of α, using the result for $\hat{\delta}_1$ from column 1, of 0.784. This implies that some 21% of workers need to find a productive partner with whom they can communicate from the region's entire population, i.e. can not match language assortatively. These individuals increase their probability to work in a communication-intensive sector by expanding the set of languages they speak.

Columns 2 to 5 show that this effect is not simply a proxy for other heterogeneity measures associated with quite different interpretations, i.e. it does not just proxy for the role of being foreigner, being of the same nationality, or having a common mother-tongue. In column 2 I control for the simple share of foreign nationals in a district at the start of the period and in column 3 I control for the Herfindahl-Hirschman index of heterogeneity in nationalities. The coefficient estimate increases slightly in both cases and both coefficients remain significant at the 5% level. In column 4 and 5 I control for the probability of a successful random match if individuals would continue speaking *only* the language they spoke primarily with their mother[70] (column 4) or *only* the language(s) they spoke primarily with their parents. Both of these measures account not only for heterogeneity induced by foreign nationals, but as well heterogeneity induced by French nationals speaking a minority language. In both cases the point estimate changes remarkably little and the point estimate relating to the alternative measure is virtually 0. However, the coefficient estimate of interest is no longer significant. The imprecision in the estimate is likely to be caused by the strong correlation of the the control variable with the main explanatory variable of interest, the conditional matching probability. (The Pearson correlation coefficient is 0.831 and 0.835, respectively.)

All of the regressions in column 1 through 5 are Ordinary Least Squares regressions. Naturally, one should be concerned that the conditional matching probability is itself determined by economic incentives which in turn might be shaped by the sectorial composition of the labor force. In fact, I make precisely this argument in sections 2.3 and 2.4. However, the analysis suggests as well a possible instrumen-

[70]This is the same as a Herfindahl-Hirschman index of mother-tongue heterogeneity.

tal variables strategy: In particular, it shows that the matching probability after language learning took place, $s_d^*(t)'\mathbf{1}(G_d'G_d)s_d^*(t)$, is correlated with the matching probability implied by the set of languages inherited from the parental generation, $s_d^0(t)'\mathbf{1}(G_d'G_d)s_d^0(t)$. For this to be a valid instrument, it needs to be uncorrelated with contemporaneous economic incentive for language learning. This will hold in the data to the extend that parents are non-strategic in the choice of the language they transmit to their children. Column 6 shows results from the corresponding instrumental variables regression. The size of the estimated effect of the matching probability (-0.120, s.e.= 0.050) remains almost unchanged, and it remains significant at the 90% confidence level. (The coefficient of interest in the first stage regression is 0.405 with standard errors equal to 0.026.)

While none of these specifications might be ultimately convincing, the results are stable across specification. The results are consistent with the interpretation that the probability of two individuals having a language in common is important for the sectorial composition of individuals joining the labor force (and do not reject Testable Implication 1).

The model implied that the share of workers joining the communication intensive sector would be 1 in case that at workers can communicate with each other, as long as the utility from working in the communication intensive sector is higher than in the agricultural sector. This will be true as long as the agricultural share is not too low, which I assume for the moment.[71] This corresponds to the hypothesis that $\delta_0 + \delta_1 = 0$ in (7) or Testable Implication 2. A t-test on the estimation results from column 1 in Table 19 fails to reject this hypothesis at conventional significance levels (t-stat= 1.175, p-value= 0.253).

Equation (10) makes precise how the sectorial composition is heterogeneous across language groups within the same district. Table 20 presents a number of regression specifications testing this implication. The dependent variable is the language group specific share of manufacturing workers. First, column 1 presents results from a simple Ordinary Least Squares regression on the probability that a worker of this group has a language in common with a randomly chosen individual from the départment. The coefficient estimate of ζ_1 is 0.087 (s.e.=0.022), significant at the 99% level. The specifications in column 2 and 3 control for a set of region and départment fixed effects, respectively. The coefficient estimate increases slightly to 0.100 (s.e.= 0.022) and 0.097 (s.e.= 0.022), suggesting that the result is not driven by any region or departement specific characteristics.

The analysis in section 2.2 suggests that the variance of the error term is inversely proportional to s_{dj}. Hence the generalised least square estimator which takes this information into account is the efficient estimation procedure for equation (10).

[71] If this assumption does not hold in the data, the power of a t-test is higher, i.e. performing a t-test is more powerful than we think.

Column 4 through 6 show equivalent specifications using Generalised Least Squares, where the variance of the error is modeled to be inversely related to the number of observations in the language group. The size of the coefficient estimate increases across specifications to in between 0.117 (when including departement fixed effects) and 0.124 (when including no fixed effects) and it remains significant at the 99% level throughout. Equation (10) suggests that $1-\hat{\zeta}_1$ is an alternative estimate of α. Taking the last estimate, this implies $\hat{\alpha}=0.883$.

Lastly, Figure 9 presents results which help to understand whether the simple model is sufficient to explain the role of language knowledge for the sectorial composition across groups. In particular, it presents coefficient estimates of ζ_1 and 90% confidence intervals from 5 separate regressions. Each of them is equivalent to column 1 of Table 20, but uses data on the districts' 2nd, 3rd, 4th, 5th and 6th biggest language group only, respectively. The coefficient estimates are in the same ball park around 0.2 (corresponding to $\alpha=0.8$).

All coefficient estimates for α – both from the group level and regional level data – are in the $[0,1]$ interval, as suggested by the theory, and range from around 0.5 to around 0.8. Hence this simple one-parameter model of language heterogeneity hence seems to have some traction to explain differences in the sectorial composition of the workforce both across départements and across language groups within départements.

2.5.3 Language Learning and Wages Differentials

I now present estimation results for λ. Table 21 presents results from estimating equation (19). Column 1 presents results from a simple Ordinary Least Squares regression for reference. The explanatory variable uses information on π_d, which measures the differential utility gain from working in the communication intensive sector. Equation (27) suggests how to calculate π_d using information on the agricultural and manufacturing wages, as well as information of η. I calculate π_d for each département.[72] As discussed in section 2.3.2, we have all reasons to belief that πs^*_{d2} is endogenous in this regression. The theory suggests to use s^0_{d2} as instrument for πs^*_{d2}. Column 2 presents results from such an instrumental variable regression. The corresponding first stage regression is shown in column 3. The excluded instrument is highly significant in the first stage. The coefficient of interest in the instrumental variable regression is negative (Testable Implication 4) and significant at the 90% level (-0.050, s.e.=0.022). Recall that $\theta_1=-\lambda(1-\alpha)$. Based on our estimate of α of 0.784 (see column 1 of Table 1) this implies the estimate of λ to be 0.231. The size of this coefficient needs to be evaluated relative to π. The mean of π is 60.8 (s.d.= 29.5). Hence the mean utility cost of learning another language ($1/\hat{\lambda}=4.3$) is

[72] See section 2.6.1 for information on how I calibrate η.

estimated to be 7% of the utility gain from working in the communication intensive sector. This estimate naturally needs to be interpreted in the context of the explicit French language learning policy, which essentially reduced the individual language learning cost. The estimate is a lower bound for the mean disutility of language learning in the absence of such a policy.

Notably, the coefficient estimate for $\log s^0_{d1}$ is not significantly different from 1 (Testable Implication 5). This can be interpreted as changes in the language group's share being only depend on the other language groups' sizes, not the own group's size.

2.6 Model Evaluation

The coefficient estimates of α and λ obtained before can be used to simulate the general equilibrium model. Obviously, the model has more unknown parameters: β and η, as well as the full set of Γ_d and γ_d. I calibrate these using the model's equations of state. I then use these together with the estimates of α and λ to simulate the stock of language knowledge and sectorial composition in 1946. To evaluate the model I compare the model's predictions against the historical development.

2.6.1 Calibration

All of a_d, s_d, G_d, w^m_d, SF_d, e and the share of expenditure on agricultural products are observable. The equations of state presented in section 2.4.1 imply how to estimate the unknown parameters η, β, Γ^m_d and γ^m_d. I assume that these parameters remain unchanged over time. Take consumption and work to be measured over a week, and working time and wages over days.

Equation (21) states that β is the expenditure on agricultural goods in units of manufacturing goods. Note that we can normalize p^m_d (which by free trade in manufacturing goods is constant across regions) to 1 by a suitable choice of units of measurement of m. The parameter β can be calibrated from a simple aggregate statistic on prices and the consumption share of agriculture.[73] I do not have data on

[73]Obviously this relies on the utility function specifying correctly preferences. We could test this, as (21) has two (closely linked) testable predictions: (i) For constant prices the amount spend on agricultural products is constant, and hence, in particular, independent of income. (ii) Changes in how much is spend on agricultural products should be predicted by changes in the relative price alone, and not, for example, total income. We can test these with cross-sectional data (in case that prices are constant across the data) and/or time-series data. I only have data on the consumption behaviour for the period from 1950 (Source: *Consommation*). Unfortunately the data published each year varies, and I cannot extract time-series data. In any case, it gives data for 1958 for 16 different social groups. The amount spend on agricultural and non-agricultural products by each group is presented in Figure 11. This corresponds to test (i) mentioned above. This seems to be supportive of the idea that a constant β (in terms of manufacturing goods) is spend on agricultural goods. If one runs a regression, the slope coefficient is still significant. However, this might partly be because the prices which those different groups are exposed to are different[74], in which case the model does not imply that the expenditure on agricultural goods is the same.

the agricultural consumption share or prices in 1930. I use data from 1956, which I deflate to 1930 prices. The weekly consumption of agricultural goods was then 84.65Fr, which is the estimate of β. This is measured in Old Franc, as are η, γ_d^m, π, Γ_d^m, wages and c_{di}, but this will be immaterial to my analysis, since only ratios are of interest (which obviously means that a further normalisation could be done).

Equation (22) suggests how to estimate η. As $V_y(y, e^*) = 1/p_m$ in the non-subsistence case and by setting $p_m = 1$ or by using data on p_m, this implies that η can be estimated from an individual level regression of observed wages on observed work hours.[75] This would take the model very seriously. The estimate is likely to be biased, as for example 'pleasure in working' might influence both e^* and w^*. I do not have individual or département specific data on working times. I take the average working time to be 6 days (per week). From the average daily wage we can calculate our estimate $\hat{\eta} = 5.35$Fr.

Equation (26) links the equilibrium firm size to Γ_d^m. I obtained data on the département average firm size in 1926 and département average wages from the 1930 French statistical yearbook. Together with $\hat{\eta}$ I back out Γ_d^m for each département. Given my estimates of Γ_d^m, I use equation (25) to back out γ_d^m.[76]

2.6.2 Simulation

Conditional on the language endowments prior to language learning of the population born between 1906 and 1926 in each département, the parameters $(\hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{\eta})$ and the region-specific parameters $(\hat{\Gamma}_d^m, \hat{\gamma}_d^m)$, the general equilibrium of language learning, manufacturing share and income differential characterised in proposition 4 can be calculated numerically.

In order to map the data to the model, I take the 'Etude de l'Histoire Familiale', but consider only the information on the language spoken primarily with the mother (or father, if no information on the mother is available) to construct the département specific shares of single-tongue language groups. These correspond to s_d^0 in equations (15)-(17). Equation (28) maps any A_d into π. Given any π, and using the estimate of λ, Corollary 2 suggests how to find numerically equilibrium language learning decisions satisfying equations (15)-(17). Equation (8) maps these into the share of the cohort which works in agriculture/manufacturing. For this I use data on the existing agricultural share in 1926 for $A_d(t-1)$ and assume $T = 2$. I perform a grid search over values of A_d to find the general equilibrium in each department d.

Another thing is that the first two groups are agriculturalists and agricultural workers, which surely have some unreported self-consumption of their produce. The take-away is that the implication of the log-linear utility function does not look totally off.

[75]Or $1/\eta$ when run the other way around, though this is not unbiased for η, even though everybody always does it.

[76]If one had data on agricultural prices *and* productivity, one could use the equilibrium price equation (24) as a robustness check. Proxies of agricultural productivity might be output per area, rainfall or land gradient.

Figure 10 presents the results. Subfigure (a) compared the actual and the simulated change in the agricultural share from 1926 to 1946. Subfigure (b) presents the actual and simulated probability that two randomly matched individuals have a language in common for the cohort of individuals born between 1906 and 1926, i.e. roughly those who entered the labour market between 1926 and 1946. From these figures it is apparent that the model predicts a substantial fraction of the co-evolution of both language abilities and the sectorial composition between 1926 and 1946.

In order to test whether the model predicts a significant share of the variation in the data and to calculate which fraction of the variation is predicted correctly by the model, consider Table 22. All columns show Ordinary Least Squares regressions. In Panel A the dependent variable is the historical change in the agricultural share between 1926 and 1946. The main regression of interest, column 1, includes as explanatory a change in the agricultural share between 1926 and 1946 predicted by the simulation of the general equilibrium model. The coefficient of interest is positive and significant (0.179, s.e.= 0.032). Its size is significantly lower than 1, suggesting that the model over-predicts the real change in the agricultural sector. Since in all previous econometric specifications there is a concern, that there might be some omitted factor, which both drives language homogenisation and the decline of the share working in the agricultural sector, I include in column 2 the initial agricultural share in 1926. The coefficient of interest on the simulated change remains in the same ball park and is highly significant, while the coefficient on the initial agricultural share is insignificant. Further, the fraction of the explained variation in the data remains virtually unchanged (R^2 increases from 0.339 to 0.346). Lastly, when only including the initial agricultural share in 1926 in column 3, the R^2 drops to 0.270. Hence, compared to a simple model of mean reversion, the general equilibrium model of language heterogeneity, language change, and structural change explains some additional 7% of the variation in the data compared to a simple model of mean reversion. Obviously, this simple model of mean reversion might partly explain the data well, since it is correlates with the predictions of the general equilibrium model set out here - in a dèpartement with high agricultural share we expect the wage differential to be big, language learning to be strong and hence the agricultural share to fall. Therefore, 7% is a lower bound on the share of the data's variation that the model explains.

Panel B show similar results for the probability of two randomly chosen individuals in a district having a language in common. The coefficient on the model's prediction is positive and significant both when controlling for the mother-tongue Herfindahl-Hirschman index in column 2 and when not doing so in column 1. However, while the model alone predicts some 61.4% of the variation in the data, it does not add substantially over and above the predictive value of mother-tongue heterogeneity.

I interpret these results as the model explaining a substantial share of the variation in the speed of structural change across French departements, but to a lesser extend the language change.

2.7 Discussion

This paper aims to be a first step to understanding the relation of language homogenisation and structural change. While there are potentially many mechanisms which might link these two processes, this paper focuses on one particular – and I think particularly relevant – channel: It analysis the role of a common language in the formation of joint productive units. Further it derives implications for the pattern of language change induced by incentives for language learning due to higher wages in the communication intensive sector. The paper provides a simple framework to organise our thinking about these topics.

The theoretical framework also highlights the key structural parameters and how these can be estimated. Using data from early 20th century France I estimate that some 21% of workers cannot match language assortatively (or have an occupation which does not allow them to). Hence language diversity might play an important role in the formation of joint productive units. Secondly, the mean utility cost of learning another language is estimated to be 7% of the utility gain from working in the communication intensive sector versus working in the agricultural sector. The estimation procedures might raise reasonable doubts, and none of the estimates might be convincing. However, they demonstrate how an empirical investigation of the model can be guided and tightly linked to the theoretical framework and where further work would be interesting.

Lastly, I calibrate the remaining parameters of the general equilibrium model in order to simulate the model and derive prediction for the language homogenisation and structural change across French départements between 1926 and 1946. I compare these predictions with the historical development. The general framework of the joint determination of language homogenisation and structural change set out in this paper explains at least 7% of the variation in the change of the agricultural share. However, it does not explain any substantial variation in the language homogenisation. Naturally, the interpretation of these results needs to rest on the specific French context. In particular, France has had an explicit policy of language homogenisation since the French Revolution, and learning of French was not costly, and unavoidable when going to school. It might well be, that in another context economic incentives for language change are more important.

B Appendix

B.1 Proofs

Proof of Proposition 1. The 'reaction functions' are given by

$$\frac{s_j}{s_j^0} = \max\{\min\{(\bar{c} - (1-\alpha)\pi s_{-j})/\bar{c}, 1\}, 0\} \qquad (29)$$

where $-j$ refers to group 1 if $j = 2$ and vice versa. Consider case a. Conjecture an interior equilibrium. Then these solve for (14). In this, when $s_2^0 < \bar{c}/((1-\alpha)\pi)$ (and hence $s_1^0 < \bar{c}/((1-\alpha)\pi)$) the denominator and nominator are positive for $j = 1, 2$ and the denominator is necessarily smaller than the denominator. Hence the interior equilibrium exists. It is unique since no boundary equilibrium exists. This is because for $s_{-j} = 0$ we have $s_j/s_j^0 = 1$, but for $s_j = s_j^0$ equation (29) implies $s_{-j} > 0$, since $s_j^0(1-\alpha)\pi < \bar{c}$. Similarly $s_{-j} = 1$ cannot be an equilibrium. Consider case b. Again conjecture an interior equilibrium. Now (9) cannot be an equilibrium, since the denominator is positive for $j = 2$, but negative for $j = 1$, while the sign of the denominator is the same for $j = 1, 2$. However, it needs to hold that $s_1, s_2 \geq 0$. When $s_1 = 0$ equation (29) implies $s_2/s_2^0 = 1$. This implies $s_1 = 0$. Hence this is an equilibrium. It is straight-forward to check that $(s_2, s_1) = (0, s_1^0)$ is not an equilibrium. Lastly, consider case c. It is easy to check that $(s_1, s_2) = (s_1^0, 0)$ and $(s_1, s_2) = (0, s_2^0)$ are equilibria as outlined before. The stability of equilibria is easily checked by comparing the slopes of the reaction functions. ■

Proof of Proposition 2. We first show that given the language learning decisions being as postulated in proposition 2, a vector $\boldsymbol{s}^*$ exists which satisfies (15), (16) and (17) (Part 1). We then show that the language learning decisions are optimal (Part 2).

Part 1: Define the following sequence

$$\begin{aligned}
s_1(n+1) &:= s_1^0[1 - F_C((1-\alpha)\pi(s_2(n)))] \\
s_2(n+1) &:= s_2^0\left[1 - F_C((1-\alpha)\pi(s_1(n) + \sum_{m=3}^{N_d}(s_m^0 - s_m(n)))\right] \\
s_j(n+1) &:= s_j^0\left[1 - F_C((1-\alpha)\pi(s_1^0 + \sum_{\{m:m\neq j,1\}}(s_m^0 - s_m(n)))\right] \forall j = 3, 4, \ldots, N_d
\end{aligned}$$

and denote $\boldsymbol{s}(n) := [s_1(n), s_2(n), s_3(n), \ldots, s_{N_d}(n)]'$. First note that $s_j(n) \leq s_j(n-1)$ for all $j = 2, 3, \ldots, N_d$ and $s_1(n) \geq s_1(n-1)$ imply $s_j(n+1) \leq s_j(n)$ for all $j = 2, 3, \ldots, N_d$ and $s_1(n+1) \geq s_1(n)$. This is true as F_C is monotonically increasing. Secondly, consider $s_j(0) = s_j^0 \forall j = 2, 3, \ldots, N_d$ and $s_1(0) = 0$. Then as long as $F_C(z) > 0$ for any $z > 0$, together with $\alpha < 1$ and $\pi > 0$, it holds that $s_j(1) <$

$s_j^0 = s_j(0)$ for all $j = 2,3,\ldots,N_d$ and $s_1(1) > 0 = s_1(0)$. By induction, $s(n)$ is a monotonic sequence, with the first element being monotonically increasing and element $j > 1$ being monotonically decreasing.
Further it is true that $s_1(n)$ is bounded from above (by at most s_1^0) and any $s_j(n)$, $j = 2,3,\ldots,N_d$, is bounded from below (by at least 0).
As $s(n)$ is monotonic and bounded, by the monotone convergence theorem it is converging. The limit satisfies the condition for an equilibrium given by (15), (16) and (17). Hence an equilibrium exists and can be found as the limit of the sequence $s(n)$.

Part 2: It is trivial that a language learner from group $j > 2$ finds it optimal to learn the language of group 1. This is because $\forall m > 1$ it is true that $s_1^0 > s_m^0$ and nobody learns the language of group j in equilibrium. For a member of group 2 to find it optimal to learn the language of group 1 it needs to hold in equilibrium that

$$s_1^* + \sum_{m=3}^{N_d}(s_m^0 - s_m^*) > \max_{\{m:m>2\}} s_m^0$$

We know by assumption $\max_m s_m^0 = s_3^0$. As $s_m^0 - s_m^* \geq 0$ for all $m > 3$ it is sufficient to show that $s_1^* > s_3^*$. This follows from (15) and (17) as $s_1^0 > s_3^0$ and $s_2^* \leq s_2^0 < s_1^0 \leq s_1^0 + \sum_{\{m:m\neq j,1\}}(s_m^0 - s_m^*)$ and F_C is monotonically increasing.
For an individual from language group 1 to prefer learning the language of group 2 over any other language, it needs to hold that

$$s_2^* \geq \max_{\{m:m>2\}} s_m^*.$$

Suppose not, i.e. there exist an s_n^* with $n > 2$ such that $s_n^* > s_2^*$. As by assumption $s_2^0 > s_n^0$ it follows that $s_2^0 - s_2^* > s_n^0 - s_n^*$. Write (17) as

$$\begin{aligned} s_2^* &= s_2^0[1 - F_C((1-\alpha)\pi(s_1^* + (s_n^0 - s_n^*) + \sum_{\{m:m\neq 1,2,n\}}(s_m^0 - s_m^*))] \\ s_n^* &= s_n^0[1 - F_C((1-\alpha)\pi(s_1^0 + (s_2^0 - s_2^*) + \sum_{\{m:m\neq 1,2,n\}}(s_m^0 - s_m^*))]. \end{aligned}$$

But then $s_2^0 > s_n^0$ together with $s_n^0 - s_n^* < s_2^0 - s_2^*$ and $s_1^* < s_1^0$ imply $s_2^* > s_n^*$, a contradiction. ■

Proof of Corollary 1. Suppose there two groups j and h, $j,h > 2$, such that $s_j^0 > s_h^0$ and $s_j^*/s_j^0 \leq s_h^*/s_h^0$. Then as well $(s_j^* - s_j^0)/s_j^0 \leq (s_h^* - s_h^0)/s_h^0$. Rewrite (17) as

$$(s_j^0 - s_j^*)/s_j^0 = F_C((1-\alpha)\pi(s_1^0 - (s_h^* - s_h^0)/s_h^0 \cdot s_h^0) + \sum_{m\neq 1,j,h}(s_m^0 - s_m^*)]$$

and similarly for group h. These then immediately imply $(s_j^* - s_j^0)/s_j^0 > (s_h^* - s_h^0)/s_h^0$, a contradiction. The proof for groups 1 and 2 proceeds similarly. ■

Proof of Proposition 3. First note that given any $\boldsymbol{s}(n)$ as defined in Corollary 2, and $\pi > \pi'$ and $\boldsymbol{s}(n)(\pi) \leq \boldsymbol{s}(n)(\pi')$, it will be true that $\boldsymbol{s}(n+1)(\pi) \leq \boldsymbol{s}(n+1)(\pi')$. Further note that $\boldsymbol{s}(0)(\pi) = \boldsymbol{s}(0)(\pi') = [0, s_2^0, s_3^0, \ldots, s_{N_d}^0]'$ implies $\boldsymbol{s}(1)(\pi) \leq \boldsymbol{s}(1)(\pi')$. Secondly, let $\boldsymbol{s}_d^*(\pi')$ denote the limit of $\boldsymbol{s}(n)(\pi')$. Consider $\boldsymbol{s}_d^*(\pi', \alpha)$. Given π this is not an equilibrium, as at least one of (15), (16) and (17) would not be satisfied if some elements $s_{dj}^*(\pi', \alpha) \in (0, s_{dj}^0)$ and F_C is strictly increasing over the relevant range. As $\boldsymbol{s}(n)(\pi)$ is monotonically below $\boldsymbol{s}(n)(\pi')$ and converges to a different limit, for the limit it needs to hold that $\boldsymbol{s}_d^*(\pi, \alpha) \leq \boldsymbol{s}_d^*(\pi', \alpha)$. The proof for the α result is similar. ■

Proof of Proposition 4. We know that $V_d^m - V_d^f$ is increasing in A_d by (28). Further we know by Proposition 3 that $\boldsymbol{s}_d$ is monotonically decreasing in $V_d^m - V_d^f$ which by the (6) implies that A_d is monotonically increasing. Hence there is a unique A_d^* such that (6) and (28) are simultaneously satisfied. ■

B.2 Tables

TABLE 18: SUMMARY STATISTICS

	(1) Mean	(2) Std. Dev.	(3) Min.	(4) Max.
Aggregate Matching Probability '06-'26	0.897	0.071	0.617	1.000
Share Foreigners 1921	0.032	0.043	0.001	0.259
Hirschman-Herfindahl Nationalities 1926	0.692	0.164	0.247	0.964
Single Mother Tongue M. Prob. '06'26	0.913	0.086	0.491	0.998
Parent Tongue Matching Prob. '06-'26	0.703	0.156	0.278	0.964
Share Agriculture 1926	0.482	0.175	0.005	0.759
Share Agriculture 1946	0.456	0.158	0.007	0.764
s^*_{d1}	0.720	0.204	0.133	0.974
s^0_{d1}	0.783	0.144	0.467	0.98
s^*_{d2}	0.057	0.087	0.000	0.408
s^0_{d2}	0.130	0.135	0.011	0.475
π_d	61.7	30.4	-7.8	177.6
Wage Agricultural Workers 1930 (Fr./day)	22.3	4.2	10.0	40.0
Wage Industrial Workers 1930 (Fr./day)	34.6	4.8	27.0	53.8
Average Number of Employees Industry 1926	2.8	1.4	1.4	9.4
π_d	60.8	29.5	-7.8	177.6

Notes: The table presents summary statistics for variables used in Tables 19 through 22. Column 1 presents the mean, column 2 the standard deviation, column 3 the minimum and column 4 the maximum. For details on the variables' definitions and data sources see Table 23.

TABLE 19: LANGUAGE KNOWLEDGE AND SECTORIAL COMPOSITION: AGGREGATE LEVEL

	(1)	(2)	(3) (OLS)	(4)	(5)	(6) (IV)
			Share Agriculture 1946, (A_d)			
Aggregate Matching P. '06-'26	-0.108**	-0.129**	-0.135**	-0.088	-0.074	-0.120**
	(0.049)	(0.048)	(0.049)	(0.083)	(0.086)	(0.050)
Share of Foreigners 1921		-0.126				
		(0.085)				
Hirschman-Herfindahl Nat. 1926			0.076			
			(0.045)			
Single Mother Tongue M. P. '06-'26				-0.011		
				(0.040)		
Parent Tongue Matching P. '06-'26					-0.018	
					(0.042)	
Share Agriculture 1926	0.882***	0.866***	0.859***	0.880***	0.878***	0.883***
	(0.030)	(0.040)	(0.040)	(0.028)	(0.028)	(0.022)
Constant	0.131***	0.162***	0.097**	0.121**	0.116**	0.142***
	(0.040)	(0.046)	(0.041)	(0.053)	(0.054)	(0.042)
R^2	0.957	0.958	0.958	0.957	0.958	0.957
N	83	83	83	83	83	83

Notes: Columns 1 through 5 present results from simple Ordinary Least Squares regressions at the departement level; column 6 presents results an instrumental variables regression. Standard errors are given in parentheses. The dependent variable in all regressions in the departement level share of the active population working in agriculture in 1946. The main explanatory variable of interest is the department level probability that any two randomly matched individuals have a language in common. Notably this measure accounts for knowledge of several languages. In column 6 I instrument for this measure with the probability that any two randomly chosen individuals have a language in common - when using only language knowledge with was transferred by the parents, i.e. not considering language abilities which have been acquired later in life. In the first stage the coefficient on the excluded instrument is 0.405 (s.e.= 0.026). The standard errors are clustered on regional level (not département) to account for spacial correlation.

TABLE 20: LANGUAGE KNOWLEDGE AND SECTORIAL COMPOSITION: GROUP LEVEL

	(1)	(2) (OLS)	(3)	(4)	(5) (GLS)	(6)
			Share Manufacturing, ($1 - A_{dj}$)			
Group Matching Probability '06-'26	0.087***	0.100***	0.097***	0.124***	0.121***	0.117***
	(0.022)	(0.022)	(0.022)	(0.018)	(0.017)	(0.015)
FE	-	Region	District	-	Region	District
R^2	0.013	0.036	0.090	0.031	0.259	0.442
N	1495	1495	1495	1495	1495	1495

Notes: Column 1 through 3 present results from Ordinary Least Squares regressions. Column 4 through 6 present results from Generalised Least Squares regressions. Standard errors are given in parentheses. The dependent variable in all regressions is the dépatement specific share of members of a language group who do work, but not in the agricultural sector. The explanatory variable in all regressions is the probability that a member of the language group has a language in common with a randomly drawn individual from the dépatement. In columns 4 through 6 these the variance of the error term is taken to be inversely proportional to the language groups size. In column 2 and 5 a full set of 22 region fixed effects are included, in columns 3 and 6 a full set of dépatement fixed effects are included. All regressions also include a constant term. The coefficients are not reported for expositional clarity. The standard errors in columns 1 through 3 are calculated using the Huber-White correction to account for potential heteroscedasticity.

TABLE 21: LANGUAGE LEARNING COST

	(1) (OLS)	(2) (IV)
PANEL A:	$\log s^*_{d1}$	
πs^*_{d2}	-0.009	-0.050**
	(0.008)	(0.023)
s^0_{d1}	1.456***	0.926***
	(0.127)	(0.304)
N	76	76
PANEL B: (First Stage)		$\pi_d s^*_{d2}$
s^0_{d2}		25.133***
		(6.392)
s^0_{d1}		0.745
		(3.611)
N		76

Notes: The table presents regression results, standard errors are given in parentheses. In Panel A column 1 presents results from an Ordinary Least Squares regression, and column 2 present results from an instrumental variable regression. The regressions in Panel A do not include a constant term; see equation (18). Column 2 in Panel B presents the corresponding first stage regression, where s^0_{d2} is used as excluded instrument, as suggested by equation (19). The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity. For definitions of the variables see Table 23.

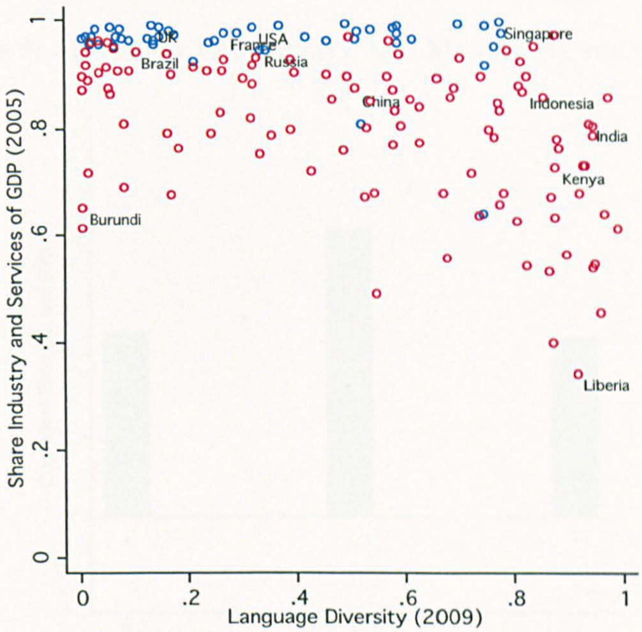
TABLE 22: MODEL EVALUATION

	(1)	(2)	(3)
PANEL A:	*Change in Agricultural Share '26-'46*		
Simulated Change Agri. Share '26-'46	0.179***	0.143***	
	(0.032)	(0.052)	
Agricultural Share '26		-0.032	-0.114***
		(0.035)	(0.023)
Constant	-0.017***	-0.003	0.031**
	(0.004)	(0.017)	(0.012)
R^2	0.339	0.346	0.270
N	78	78	78
PANEL B:	*Aggregate M. Prob. '06-'26 Cohort*		
Simulated Matching Prob. '06-'26 Cohort	0.377***	0.115*	
	(0.040)	(0.066)	
Single Mother Tongue M. Prob. '06-'26 Cohort		0.238***	0.318***
		(0.050)	(0.027)
Constant	0.625***	0.652***	0.680***
	(0.033)	(0.030)	(0.022)
R^2	0.614	0.699	0.686
N	78	78	78

Notes: The table presents results from Ordinary Least Squares regressions at the department level. Standard errors are given in parentheses. In Panel A the dependent variable is the change in the share of the working population in agriculture between 1926 and 1946. In Panel B the dependent variable is the Aggregate Matching Probability amongst the cohort of individuals born between 1906 and 1926. The explanatory variables 'Simulated Change Agri. Share Õ26-Õ46' and 'Simulated Matching Prob. Õ06-Õ26 Cohort' are predicted values from the simulation of the general equilibrium model as described in section 2.6.2. For definitions of the other variables see Table 23. The standard errors are calculated using the Huber-White correction to account for potential heteroscedasticity.

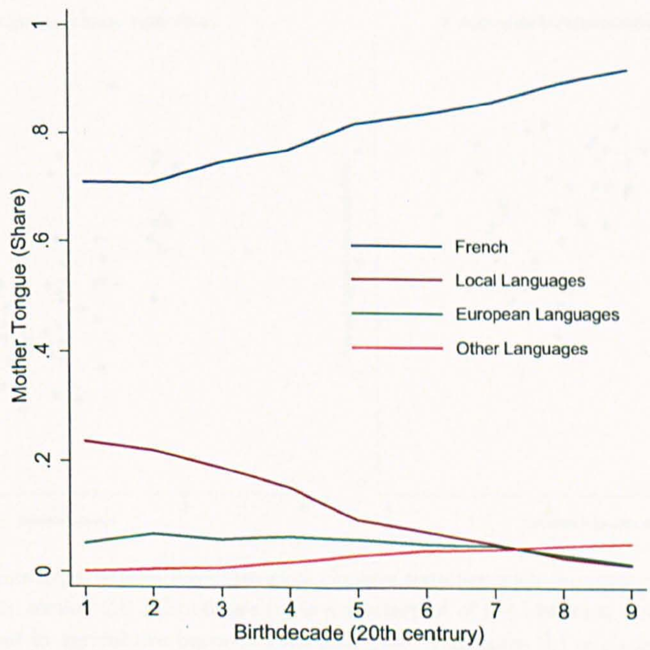
B.3 Figures

FIGURE 7: LANGUAGE DIVERSITY AND SECTORIAL COMPOSITION



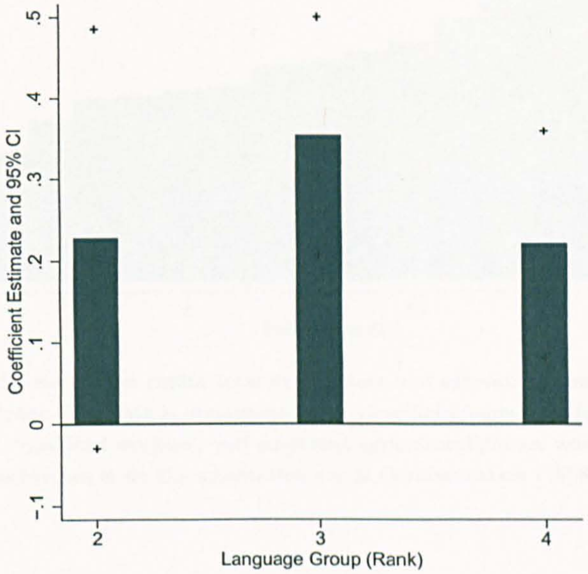
Notes: The figure is a scatterplot of national level data on language diversity in 2009 (Source: Ethnologue 2009), a standard Hirschman-Herfindahl index of mother-tongue diversity, and the share of industry and services in 2005 GDP (Source: World Development Indicators 2005). Countries with GDP below the average per capita GDP (in 2005 PPP, corresponding to 11.588$) are depicted in red, other countries in blue. Selected countries are labelled. The raw correlation between the GDP measure and language diversity is -0.259, and the correlation between the industry and services share in GDP and the diversity index is -0.464. (N=158)

FIGURE 8: 20TH CENTURY MOTHER TONGUES IN FRANCE



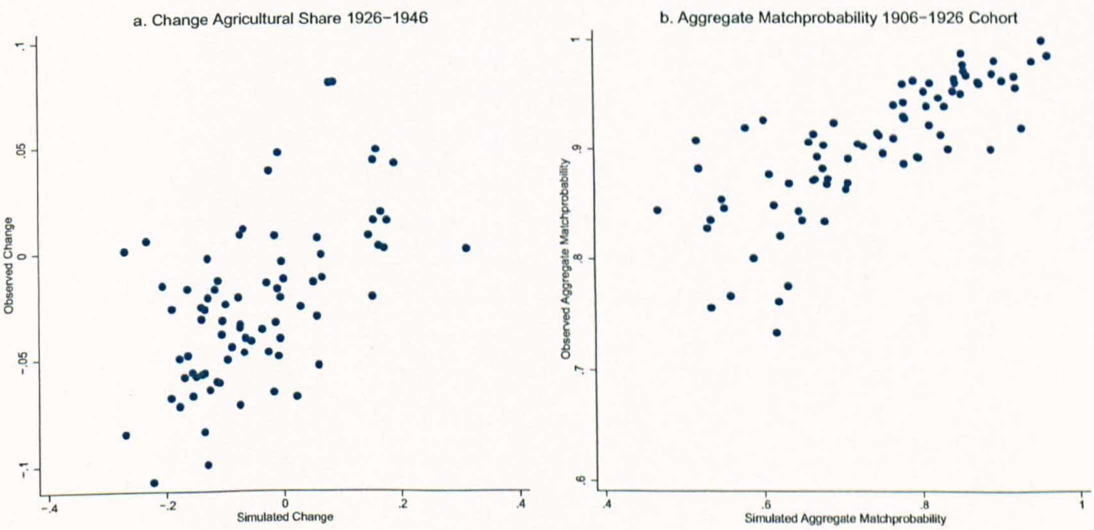
Notes: The figure presents data on the share of mother tongues – defined as the language mainly spoken with the mother (or father, if no data on the mother is available) – in France by birth cohort. The first cohort was born between 1900 and 1909, the second cohort from 1910 until 1919 and so on. The French local languages which are recognised in the database are 'français', 'alsacien', 'breton', 'catalan', 'platt moselan', 'basque', 'corse', 'francontois', 'franco provençal', 'flamand langues d'Occ', ' langues d'Oil', 'franco-provencal', 'créoles à base française'. (Source: Institut National de la Statistique et des Études Économiques (1999))

FIGURE 9: MATCHING PROBABILITY AND MANUFACTURING WORK



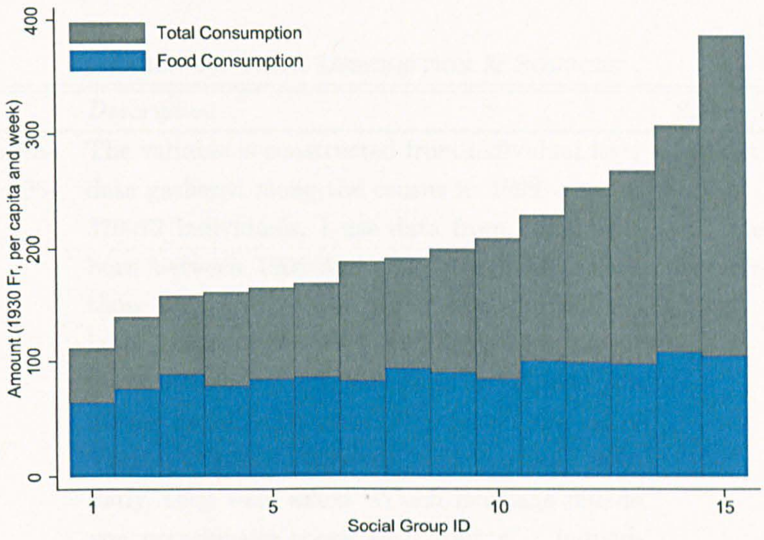
Notes: The figure presents regression coefficient estimates and confidence intervals. The estimates are the coefficients from three separate group level regressions of the 'Share in Manufacturing' on the 'Group Matching Probability '06-'26'. These correspond to the regression shown in column 1 of Table 20. However, while those results use the pooled data for all language groups, here I run separate regressions for the 2nd, 3rd or 4th biggest language groups, respectively.

FIGURE 10: SIMULATION VERSUS HISTORICAL CHANGE



Notes: The figures presents département level historical changes together with the predictions from the simulation of the model as outlines in section 2.6.2. Subfigure (a) is a scatterplot of the historical and simulated change of the share of workers employed in agriculture between 1926 and 1946. Subfigure (b) is a scatterplot of the simulated probability of any two individuals having a language in common and the observed aggregate matching probability. For the variable definitions and data sources see Appendix B.4.

FIGURE 11: FOOD CONSUMPTION SHARE, 1956 (BY SOCIAL GROUPS)



Notes: The figure present the weekly per capita total expenditure and expenditure on food across France in 1956. All amounts are in French Franc. The data is disaggregated by 15 social groups. The first 6 groups are, from left to right, 'agricultural workers', 'unskilled workers', 'self-employed agricultural','house workers', 'not active', 'workers'. For details see Centre de Recherches et de Documentation sur la Consommation (1958).

B.4 Data Description and Sources

TABLE 23: DATA DESCRIPTION & SOURCES

Variable	*Description*	*Source*
Aggregate Matching Probability '06-'26	The variable is constructed from individual level data gathered along the census in 1999. Out of 370072 individuals, I use data from individuals born between 1906 and 1926 (N=32606), since those are roughly those individuals who join the labor force between 1926 and 1946. The respondents were asked: 'Which language did/do you mainly speak with your r?' where r are all of 'mother', 'father', 'children', and 'partner'. Similarly, they were asked 'Which language did/do you occasionally speak with your r?' Individuals could give several answers to each question. For each individual I construct a vector g_{di} (with the length of all languages which appear in the data) with element equal to 1 if the respondent answers to have spoken that language frequently or occasionally with someone. From these I calculate all existing language groups and their départment specific share. Then I compute $s_d(t)'\mathbf{1}(G_d{}'G_d)s_d(t)$.	Institut National de la Statistique et des Études Économiques (1999)
Single Mother Tongue M. Prob. '06'26	This measure is calculated analogously to 'Aggregate Matching Probability '06-'26'. The only difference is, that in the vector g_{di} only one element is set to 1, which is the language which is reported to have been spoken mainly with the mother. If several languages are reported, only the first one is considered only. If no information on the main language spoken with the mother is available, information on the language spoken with the father is used, if available.	Institut National de la Statistique et des Études Économiques (1999)
Parent Tongue Matching Prob. '06-'26	This measure is calculated analogously to 'Aggregate Matching Probability '06-'26'. The only difference is, that in the vector g_{di} only element are set to 1, where the individual reports that he used mainly this language when talking to either his mother or his father.	Institut National de la Statistique et des Études Économiques (1999)
Share Foreigners 1921	This is the departemental share of foreigners in 1921.	Institut National de la Statistique et des Études Économiques (1949, pp. 355-356)

CONTINUED ON NEXT PAGE

TABLE 23 – CONTINUED FROM PREVIOUS PAGE

Variable	*Description*	*Source*
Hirschman-Herfindahl Nationalities 1926	I use departemental data on the number of individuals with non-French nationality, for the 30 nationalities reported in the 1926 census (of which one is 'other nationalities') together with data on the departemental population in 1926 (reported in the 1968 census). I calculate the share of each nationality (including French) and compute the departemental Hischman-Herfindahl index.	Statistique Générale de la France (1931b, pp. 230-233), Institut National de la Statistique et des Études Économiques (1968, pp. 1152-1155)
Share Agriculture 1926, 1946	Departemental share of the active working population which is working in agriculture, forestry and fishery.	Statistique Générale de la France (1931a, pp. 145-146) Statistique Générale de la France (1951, p. 73)
Share Manufacturing	This measures the fraction of the respondents in each départment and language group which work, but do not work in agriculture.	Institut National de la Statistique et des Études Économiques (1999)
Group Matching Probability Õ06-Õ26	This variable is calculated similarly to 'Aggregate Matching Probability '06-'26', however in the last step the départment-language group specific $\mathbf{1}(g_{dj}{}'G_d)s_d(t)$ is calculated.	Institut National de la Statistique et des Études Économiques (1999)
Wage Agricultural Workers 1930 (Fr./day)	Departemental average daily wage of male agricultural workers in 1930.	Statistique Générale de la France (1931a, p. 249)
Wage Industrial Workers 1930 (Fr./day)	Departemental average daily wage of male workers in the departmental capital in 1930. Where no data for 1930 is available, I used inflation adjusted (7.382%) wage data from 1929.	Statistique Générale de la France (1931a, p. 248)

CONTINUED ON NEXT PAGE

TABLE 23 – CONTINUED FROM PREVIOUS PAGE

Variable	*Description*	*Source*
Average Number of Employees Industry 1926	The statistical yearbook reports data on the France-wide number of firms in different size categories (no employee, 1 to 5, 6 to 20, 21 to 100 and more than 100 employees), split up by sectors. It as well reports the number of persons working in firms of a certain size and sector. I uses this data to construct the average number of workers in non-agricultural (agriculture, fishery, forestry) firms, for each category. For example, non-agricultural firms with 1-5 employees have on average 1.787 workers. I then use departemental data on the distribution of firms across these categories (for example, I know that in Ain 24150 firms exist with 1 to 5 employees) together with the average number of workers in non-agricultural firms in this category to calculate the average non-agricultural firm size.	Statistique Générale de la France (1931a, pp. 144, 147)
π_d	This is calculated using equation (27), where p_m is normalised to 1, η is calibrated in section 2.6.1, and 'Wage Industrial Workers 1930 (Fr./day)' and 'Wage Agricultural Workers 1930 (Fr./day)' are used for w_d^m and p_d^f/γ_d^f.	*own calculations*
s_{d1}^0, s_{d2}^0	This data is constructed like 'Parent Tongue M. Prob. '06'26', however the data for the size of each language group is preserved in this format (i.e. not aggregated to a departemental index). The subindices refer to the biggest and second-biggest language groups.	Institut National de la Statistique et des Études Économiques (1999)
s_{d1}^* , s_{d2}^*	This data is constructed like s_{d1}^0, s_{d2}^0, however not only languages spoken usually with the parents, but also languages usually spoken with the children and friends are considered. The subindices refer to the same language groups as those of s_{d1}^0, s_{d2}^0.	Institut National de la Statistique et des Études Économiques (1999)

3 Out Of Your Mind: Eliciting Individual Reasoning in One Shot Games[77]

3.1 Introduction

Equilibrium concepts have proven to be a powerful tool in economics, political science, international relations and other fields when trying to understand and predict strategic behaviour of individuals, firms, countries, and other entities. However, manifold experimental studies of human behaviour have shown that the equilibrium concept does poorly in predicting outcomes of one shot games, even when a unique equilibrium exists (see for example Nagel, 1995, Rubinstein, Tversky, and Heller, 1996 or Camerer, 2003 for an overview). This class of games reflects *strategic situations without precedent* which are faced, for example, by consumers who – otherwise price-takers – buy a house and bargain with the seller about its price.

A leading explanation for the failure of equilibrium concepts to explain and predict behaviour in one shot games is that players may not believe that other players choose an equilibrium strategy. This explanation appears particularly pertinent in unprecedented strategic situations, where learning cannot cause a convergence of beliefs and strategies to equilibrium.[78] The level-k model of reasoning, as first proposed by Nagel (1995) and Stahl and Wilson (1995), postulates the following alternative belief structure: There exist so-called level-0 players, who do not play strategically. The model defines level-1 players to best respond to what they believe level-0 players do. Level-2 players form a belief about the fractions and strategies of lower level players and best respond to this. This process continues for higher level players. Hence the model assumes a hierarchy of types who best respond to non-equilibrium beliefs, referred to as *levels*. They distinguish themselves by the number of iterated best responses to the distribution of level-0 actions.[79]

Although there is extensive empirical work in support of the level-k model, little is known about its anchoring elements: Do non-strategic *level-0* players exist? How do they choose their actions? And do strategic players correctly anticipate the actions of non-strategic players? Answers to these questions are required for the

[77] The work in this chapter was carried out jointly with equal share by Stefan P. Penczynski and me.

[78] A distinct explanation could be that players fail to choose the best response to their belief about the other players' strategies. This is the idea behind the Quantal Response Equilibrium (QRE) proposed by McKelvey and Palfrey (1995) in which actions of higher expected payoff are more likely.

[79] Applications of the level-k model include dominance solvable games (Nagel, 1995; Costa-Gomes and Crawford, 2006), normal-form games (Stahl and Wilson, 1995; Costa-Gomes, Crawford, and Broseta, 2001), two-person zero-sum games with non-neutral framing (Crawford and Iriberri, 2007a), common-value auctions (Crawford and Iriberri, 2007b), and coordination games (Crawford, Gneezy, and Rottenstreich, 2008). For a more complete overview see Camerer, Ho, and Chong (2004) and Crawford, Costa-Gomes, and Iriberri (2010).

model to have predictive power.

The lack of empirical evidence is no coincidence. Firstly, the prominent theoretical papers typically assume level-0 actions to be uniformly distributed over the action space and the beliefs of higher level players to be non-heterogeneous and consistent with the actions of level-0 players. These assumptions naturally exclude any investigation of the level-0 actions and beliefs. We generalise standard versions of the level-k model by allowing for (i) a non-uniform level-0 action distribution, (ii) heterogeneous level-0 beliefs and (iii) the distribution of level-0 beliefs to be independent of the distribution of level-0 actions. These are necessary generalisations to answer the questions of interest. To fix ideas, we present a formal version of this generalised model in Section 3.2.

Secondly, there are important empirical obstacles to investigating these questions: choice data alone can typically be explained by various reasoning patterns – different levels of reasoning and associated level-0 beliefs or actions – and hence are not informative. In particular, non-strategic behaviour cannot be robustly identified and the level-0 beliefs are difficult to uncover. Moreover, in repeated games individuals are undergoing a learning process, which renders the use of a sequence of observations difficult.

We present two methods which allow to investigate the above questions. Both rely on an experimental design that give access to *incentivised* written accounts of individual reasoning which are stated at the time of the decision-making. These accounts are obtained through a particular team communication protocol which can be applied to any one shot game. We played a standard 'beauty contest' game: each participant is asked to state a number between 0 and 100 and the participant whose number is closest to 2/3 of the average of all numbers wins a prize. The unique Nash equilibrium of this game is to play 0 for all participants.

The first empirical strategy is to obtain a non-parametric estimate of the distribution of level-0 beliefs and level-0 actions from the written accounts of reasoning. We detect non-strategic reasoners and obtain an estimate of the distribution of level-0 actions. We find that at least 20% of the participants play non-strategically, in the sense of not attempting to best respond to any belief. The mode and median of the actions of these players are close to 60, and the mean is significantly higher than 50. This is consistent with the interpretation that salience due to the multiplier of 2/3 shapes the actions of non-strategic players. It suggests that level-0 players can be understood as having no stronger reasons for choice other than salience. Further, we analyse stated level-0 beliefs of higher level players. While the majority of them starts their reasoning at exactly 50, there is substantial heterogeneity. We cannot reject the hypothesis that the beliefs are, on average, correctly anticipating the mean of the level-0 actions.

The second empirical strategy is to estimate the structural parameters of our generalised level-k model with a maximum likelihood estimation. In the estimation we make use of information about the players' sophistication from the written accounts. We estimate about one third of participants to be playing non-strategically. The distributions of level-0 actions and beliefs have an estimated mean of 58 and 54, respectively. These are close to the non-parametric estimates obtained with the first empirical strategy. The belief distribution is more concentrated than the action distribution. However, it still has substantial variance. The estimation of these structural parameters is only possible under the generalised version of the model. Therefore, acknowledging heterogeneity in the beliefs does not only seem correct, but it as well enables novel estimations of the level-k model's structural parameters.

The two empirical strategies provide consistent evidence on the questions this paper seeks to answer. We show that around one third of the participants play non-strategically. The non-strategic level-0 actions are not uniformly distributed. They might be shaped by salience considerations. Higher level players correctly anticipate the non-uniform distribution of non-strategic actions.

After presenting the generalised level-k model in the next section, this paper proceeds as follows: In section 3.3 we discuss the existing empirical literature on the level-k model. We then present the experimental design and procedures in sections 3.4. Results from the two empirical strategies are presented in sections 3.5 and 3.6. The final section offers concluding comments.

3.2 Generalised Level-k Model

Consider a game in which N players simultaneously choose an action $x \in X$. Level-k type models deviate from equilibrium models in that they allow for heterogeneity in the belief about other players' actions. A player is characterised by a level of reasoning k, $k = 0, 1, 2, 3, \ldots$. Denote the fraction of level-k types in the population with l_k and define $\boldsymbol{l} := \{l_0, l_1, l_2, \ldots\}'$. Level-0 players are defined as playing non-strategically in the sense that their action does not depend on a belief over other players' actions. We denote the distribution of their actions by $g^0(x \mid \boldsymbol{\theta}^0)$, where $\boldsymbol{\theta}^0$ is a parameter vector characterising the distribution, and refer to it as 'level-0 distribution'. Players of a higher level of reasoning ($k \geq 1$) are thought to best respond to the actions of lower level players. The belief about the distribution of those actions is derived from (i) a belief about the level-0 distribution and (ii) a belief about the relative proportions of lower-level players. We refer to the belief about the level-0 distribution as 'level-0 belief'. Note that in most games, the player only needs to form a belief about one or several statistics of the level-0 distribution, in the sense that conditional on knowledge of these statistics, information about other aspects of the level-0 distribution leaves the best response unchanged. For example in the 'beauty contest' game a level-k player, $k \geq 1$, only needs to form a belief about the mean of the level-0 distribution. We allow the level-0 belief about the vector of statistics $\boldsymbol{d}$ to be heterogeneous across individuals. Let its distribution be characterised by the probability density function $g^b(\boldsymbol{d}\,|\,\boldsymbol{\theta}^b)$, where $\boldsymbol{\theta}^b$ is again a vector of parameters. Denote the belief of a level-k player with $k \geq 1$ about the proportion of level-i players in the population as $b_k(i)$, $i \leq k-1$. Given this belief a level-2 player can calculate the distribution of actions of lower level players and best respond to it. Higher level players find their best response analogously.[80] In conclusion, the strategy of a level-k player, $k \geq 1$, is found as probability distribution over

$$s_k(b_k(\cdot), g^b(\cdot)) = \underset{x \in X}{\operatorname{argmax}}\, u(x; b_k(\cdot), g^b(\cdot)). \tag{30}$$

Given an assumption on the anchoring level-0 distribution, the level-0 belief, the true level-k distribution, and a specification for $b_k(\cdot)$, the level-k model makes a probabilistic prediction about the frequencies of actions.

This general level-k model nests the models by Nagel (1995), Stahl and Wilson (1995), Costa-Gomes, Crawford, and Broseta (2001), Camerer, Ho, and Chong (2004) and Costa-Gomes and Crawford (2006) as special cases.[81] All of these models assume that the level-0 actions are uniformly distributed over the action space, that higher level players' level-0 beliefs are non-heterogeneous and that they correctly anticipate the actions of level-0 players.

[80] Note that for players with $k \geq 2$ higher-order beliefs need to be specified. In particular, a player needs to form a belief over the lower-level players' beliefs of the level-0 distribution. Similarly, a $k \geq 3$ player needs to form a belief about the population beliefs of player i, $2 \leq i \leq k-1$, i.e. a level-3 player needs to know what a level-2 player believes the relative proportions of level-0 and level-1 players are. Here and in the literature, these higher-order beliefs are assumed to be consistent with the beliefs of the lower-level player.

[81] Nagel (1995) assumes the players' population beliefs for $k > 0$ to be degenerate on $k-1$: $b_k(i) = 1$ if $i = k-1$ and $b_k(i) = 0$ otherwise. She assumes – as all later versions of the level-k model – the level-0 distribution to be uniform and every player's level-0 belief to be consistent, which makes $g^b(\cdot)$ degenerate. Stahl and Wilson (1995) study normal form games and present a version of the level-k model where best responses are calculated with error. The uniform level-0 distribution, $g^0(\cdot)$, reflects a fully imprecise best-response. Levels 1 and 2 best respond with error, playing actions with a higher expected payoff with a higher probability. Higher levels than $k = 2$ are not considered and the population distribution belief is not restricted to be degenerate. Costa-Gomes, Crawford, and Broseta (2001) and and Costa-Gomes and Crawford (2006) model the level-k types in a similar fashion as Nagel (1995), but assume level-0 players not to exist. Further they consider 'equilibrium' types that play the Nash equilibrium strategy and call 'sophisticated' types those players that best respond to the actual distribution of others' responses. Camerer, Ho, and Chong (2004) introduced the 'cognitive hierarchy' model where the players' beliefs reflect the true relative frequencies of lower level types and the true distribution of types $l(k)$ follows a Poisson distribution with parameter τ. Formally, for all $k > 0$, $b_k(i) = l(i;\tau)/\sum_{m=0}^{k-1} l(m;\tau)$, where $l(k;\tau) = \tau^k e^{-\tau}/(k!)$.

3.3 Literature

The paper relates to an extensive empirical literature on the level-k model of reasoning. Most of this literature is concerned with estimating the distribution of level-k types in a given population. For example, Camerer, Ho, and Chong (2004) use a GMM estimator, essentially choosing the parameter of a one-parameter version of the level-k model (as well referred to as 'cognitive hierarchy model') to match the mean of the action data. This is an elegant methodology to estimate the level-k distribution that assumes certain level-0 action and belief distributions. An approach which does not rely on a structural model is proposed in Bosch-Domènech, Montalvo, Nagel, and Satorra (2004), who use a large dataset to fit a set of normal distributions to action data from various 'beauty contest' games. Some of the superimposed normal distributions are then associated to underlying levels of reasoning, which provides both estimates of the type distribution and the choices for individual types.

In contrast, we estimate the various parameters of a structural model using maximum likelihood estimation in our second empirical strategy. This is typically not possible, since the level-k model makes stark choice predictions when the distribution of beliefs is assumed to be non-heterogeneous. For example, a uniform level-0 distribution in the 'beauty contest' game, together with non-heterogeneous, consistent beliefs about the mean of level-0 actions, imply a degenerate belief at 50. Any action slightly off 33, 22, etc. will be attributed to level-0 play. A maximum likelihood estimation then typically yields almost all players to be level-0 players and fits the level-0 distribution to the full-sample action distribution. We instead explicitly model heterogeneity in the level-0 belief, which causes heterogeneity in actions of players of the same level-k, $k > 0$, and allows us to separate this from level-0 play. In this sense the generalisation of the model is key to our estimation strategy.

A central challenge faced by the empirical literature is that the level-k model can typically explain a given *action* in a one-shot game with several reasoning patterns.[82] Like we do, other studies try to circumvent this problem by obtaining additional information which allows to conclude about the underlying reasoning. Costa-Gomes, Crawford, and Broseta (2001) and Costa-Gomes and Crawford (2006) obtain for each player multiple choices made in subsequently played variants of a game without feedback. Under the assumption of a constant reasoning level, they then match a player's sequence of choices to a typical 'fingerprint' of, say, a level-1 player. This provides estimates of individual levels of reasoning. In addition to the 'fingerprint',

[82] For example, a player who chooses 33 in the standard 'beauty-contest' game might do so because he is a level-1 reasoner who believes that level-0 reasoners play on average 50. But he might just was well be a level-0 reasoner who has chosen the number at random or a level-2 reasoner who beliefs that the population is composed of a combination of level-1 and level-0 reasoners who on average choose 50.

they used information search data to identify types of players.[83] Tracking both choices and response times, Rubinstein (2007) associates longer response times with more cognitive effort, differentiating between cognitive, instinctive and reasonless choices. In a similar vein, Agranov, Caplin, and Tergiman (2010) incentivise and observe provisional choices over time in order to get insights in the choice process, including initial naive considerations. In her original study, Nagel (1994) asked the participants after the experiment to verbally state the reason for their chosen action and then classified and analysed comments of participants. A descriptive analysis of optionally given verbal comments received in the context of a newspaper experiment is presented by Bosch-Domènech, Montalvo, Nagel, and Satorra (2002). These studies do not analyse the non-strategic actions and beliefs thereof.

Our experimental design analyses written accounts of reasoning. We carefully ensure that these are written during the actual thinking period and provide an incentive to the participants to state their reasoning fully and clearly.[84] This is central to our empirical strategy, since only with such incentives we can be sure that the reasoning process is fully represented in the written accounts. In the absence of such incentives one is likely to underestimate the level of reasoning.

At the heart of our experimental design lies the use of team communication as a means of observing individual reasoning. The experimental literature has used team setups on various occasions to obtain insights into the reasoning process of participants and to investigate the performance of teams as opposed to individuals. In this respect, our experimental design is related to the innovative study by Cooper and Kagel (2005) who were the first to let team players communicate via an instant messenger, allowing the experimenters to observe the speed of learning in strategic play. We use a communication protocol in which the message is written prior to any team interaction, so that it purely reflects individual reasoning in a one-shot game.

3.4 Observing Individual Reasoning

We present two strategies to estimate the distribution of level-0 actions and beliefs thereof. Both empirical strategies rely on an experimental design which allows us to obtain incentivised accounts of individual reasoning. This section explains our design. It emphasises how our design ensures that individual participants have an incentive to state their reasoning fully and clearly. It as well describes how we extract information from these written accounts by classifying the reasoning pattern along the lines of a general level-k model. We corroborate the robustness of our findings

[83] This method was introduced by Camerer, Johnson, Sen, and Rymon (1993).

[84] Indeed, the standard theme in protocol analysis, the research field in psychology concerned with methods of eliciting verbal accounts from participants, is that "[t]he closest connection between thinking and verbal reports should be found when participants were instructed to focus on the task while verbalising their ongoing thoughts" (Ericsson, 2002, p. 983).

and the replicability of the classification procedure.

3.4.1 Experimental Design

In order to elicit the reasoning underlying a player's action, we designed the following game structure: Individuals are randomly assigned in teams of two players. Their payoff in the game depends on a joint 'team action'. To determine this, both players are given the chance to choose an action – which we call the 'final decision'. Then one player's decision is chosen randomly, with probability one half as the team action. Consequently, each player faces a 50% chance of having her partner's final decision determining the team action. The players hence have an incentive to ensure that their team partner's final decision is as sound as possible. Importantly, the players are given the possibility to convince their partner of the optimal team action. In particular, players are allowed to write one message to their team partner, which consists of a 'suggested decision' and a justifying text. This text is unlimited in size and its writing is not limited in time. The messages are exchanged *simultaneously* once both players have entered their message, and thereafter the players take their final decision individually.

The simultaneous exchange of a single message ensures that every explanatory statement and suggested decision in the first round is written without any previous communication with the team partner, hence reflecting an individual's reasoning. It is therefore this first message and suggested decision which we will analyse in this paper in order to understand individual decision making in situations of strategic interaction.[85] Importantly, under the reasonable assumption that the best way to convince one's team partner is by explaining him the own reasoning, this design gives an incentive to write down the reasoning process as fully and clearly as possible. It is generally applicable in one shot games and we believe it constitutes an important methodological contribution of our paper.

The design has two potential caveats. First, the suggested decision is taken while knowing that the opponents are teams of 2 players. This leads, if anything, to a population belief with more weight on higher levels due to the team reasoning. Second, the decision is taken while justifying it in a message. This might lead to a more thoughtful decision. However, we cannot exclude the possibility that the requirement to explain one's decision can cause stress, leading to a lower level of reasoning. In any case, if these distortions were present, this should show up in the action data. It is comforting that the distribution of suggested decisions is similar to other non-communication treatments in the literature, e. g. Nagel (1995).

---

[85]The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

3.4.2 Experimental Procedures

We conducted the experiment in the Experimental Economics Laboratory of the Department of Economics in Royal Holloway (University of London). In 6 sessions we played three rounds of the 'beauty-contest' game. Since in this paper we are interested in individual reasoning, we only analyse the first round suggested decision and the accompanying first message, i. e. with the activities that took place before any interaction of the team players.

At the start of each session the participants were made familiar with the structure of the experiment and the messaging system in two practice rounds. We used the same software as above for the practice rounds, but asked the teams to find the answer to two unrelated questions. Since we wanted to avoid any pre-treatment sensitisation to strategic considerations, we asked them to provide the year of two historic events. The questions in the test round were chosen to be relatively difficult to stimulate the use of the messaging system. The participants of the experiment were paid a show-up fee of £5 and the winning team won a prize of £20 (£10 per team player).

A total of 84 individuals participated in our experiment. Sessions had 12, 14 or 16 participants. The participants were mainly undergraduate students in Royal Holloway and all of them were recruited by the host institution. Out of the 84 students 15 were studying Economics, 13 of them being in their first year of studies, one in the second year, one being in the third year. 16 of the 84 students had received some form of training in game theory, but only 5 had been confronted with the 'beauty contest' game. The majority of students had participated in an economic experiment before.

3.4.3 Classification of Communication Transcripts

We use the information elicited on individual reasoning in the following way: First, we identify the cases in which no level-k type of reasoning is undertaken. Then, conditional on a level-k reasoning being applied by the player, which includes non-strategic play, we uncover two sets of data: (a) the maximum and minimum number of steps of reasoning which can be interpreted into the message, including possibly 0, and (b) the level-0 belief which a player states, if any.

In particular, two research assistants read the messages and classified the type of reasoning with the following procedure:

1. To investigate the prevalence of reasoning patterns different from level-k reasoning we asked the RAs to indicate whether the player puts forward equilibrium reasoning. For this it was not necessary that the player actually played the unique equilibrium strategy.[86] The RAs were further instructed to denote

[86] We call a player a 'sophisticated' type, if she recognised the equilibrium, but played an action

whether the player applied an iterated elimination of dominated strategies. For this it is necessary that *first* some actions are excluded and then a strategy is formed for the remaining action space.[87]

2a. If any level-k reasoning was explained in the message, we asked the RAs to indicate if the argument contained a belief about others' play that served as a starting point for best responses, but was in itself not derived by choosing a best response. If so, we asked them to denote this level-0 belief.[88]

2b. Lastly, we were interested in how many steps of reasoning the player applies. When designing the classification procedure we were worried that in some cases it might not be possible to identify from the communication *exactly* how many steps of reasoning were applied.[89] We therefore asked the classifiers to only indicate the lowest level of reasoning which is clearly stated and the highest level of reasoning which could possibly be interpreted into the messages.[90] We refer to these as 'lower bound' and 'upper bound', respectively. We instructed the classifiers to consider as level-0 a player whose message does not exhibit "any strategic reasoning whatsoever". This might arise as a result of choosing a number randomly or based on non-strategic considerations such as taste. We emphasised that for this classification to be chosen, it was important that the player was not in any way best responding to what he thought others would play.

When designing the classification procedure we intended to avoid two potential concerns: First, the classifiers might try to extract more information than the messages actually contain. We therefore instructed the classifiers to only enter information when it was clearly contained in the message.[91] Second, we were concerned that in the 'beauty-contest' game in the case of an ambiguous statement relatively low suggested decisions might lead the classifiers to indicate a higher lower-bound

different from the unique equilibrium action.

[87]For example the statement "Everybody plays on average 50 so I should not play higher than 34" is not an iterated elimination of dominated strategies.

[88]For completeness, we also asked whether an argument revealed a population belief distribution. If so, the classifiers were asked to indicate whether it was degenerate or non-degenerate.

[89]Think for example of the imaginary statement: "I presume everybody else will play 33, so let us play 22." This clearly exhibits one step of reasoning. But it seems possible, too, that the player skipped the first step of his reasoning when writing down his argument.

[90]The instructions specified that the classifiers, after writing down the lower bounds should be able to say to themselves: "It seems impossible that the players' level of reasoning is below this number!", and after writing down the upper bounds: "Although maybe not clearly communicated, this statement could be an expression of this level. If the player reasoned higher than this number, this was not expressed in the statement!"

[91]The instructions were self-contained and were not complemented by verbal comments. The instructions were written by the two authors, of whom one had taken a look at the communication transcripts beforehand. The instructions can be obtained from the authors upon request. Remaining questions were answered via an e-mail list that included all four persons involved and which can be obtained from the authors.

on the level of reasoning than was clearly exhibited. In contrast when indicating the upper-bound, knowing about low choices should, if anything, lead the classifiers to indicate a higher upper-bound. We therefore split the classification of the messages into two parts. We did not reveal the choice data to the classifiers when asking for the lower-bound but revealed it subsequently when asking for the upper bound.[92]

The classification was undertaken by two Ph.D. students in the Department of Economics at LSE. First they classified the transcripts individually. After this phase their classification of the lower bound coincided for 77% of all participants and the classification of the upper bound coincided in 76% of all cases. Then the two RAs met to reconcile their judgements and provide a joint classification, if possible. We only use data on which they could agree in the reconciliation.

Later we asked further 6 RAs to again classify the lower bounds. Table 29 in appendix C.7 shows that for 70 out of 78 messages (~90%), 6 or more of the 8 classifiers agreed on exactly one level.[93] We take this as comforting evidence that our method of classification is robust, provides informative insights about individual reasoning and can easily be replicated.

3.5 Experimental Results

3.5.1 Action Data

Table 24 presents aggregate summary statistics for all 6 sessions. Figure 12 shows histograms of the suggested decision and the final decisions aggregated over all sessions. The suggested decision is comparable to the first period's decision of other experiments with individual participants. However, the final decision is not, since the participants have, at the time of taking this decision, already received a message from their team partner. For the subsequent analysis and classification of the individual reasoning we will exclusively use the suggested decision.

The data on the suggested decision is similar to data generated in other comparable experiments in having similar means and a high fraction of choices between 20 and 50. The original study by Nagel (1995) had an average of 36.6, which is slightly lower than ours.[94] A concern with our design is that having to communicate to the team partner might increase the participants' level of reasoning, e. g. because the participants would examine the task at hand more thoroughly in order to state sensible arguments in the communication. The fact that our data exhibits – if anything

[92]Other studies applying a classification use a similar procedure in order to avoid any unconscious alignment of the classification with the choice data that might result from implicit assumptions (for example Rydval, Ortmann, and Ostatnicky, 2009).

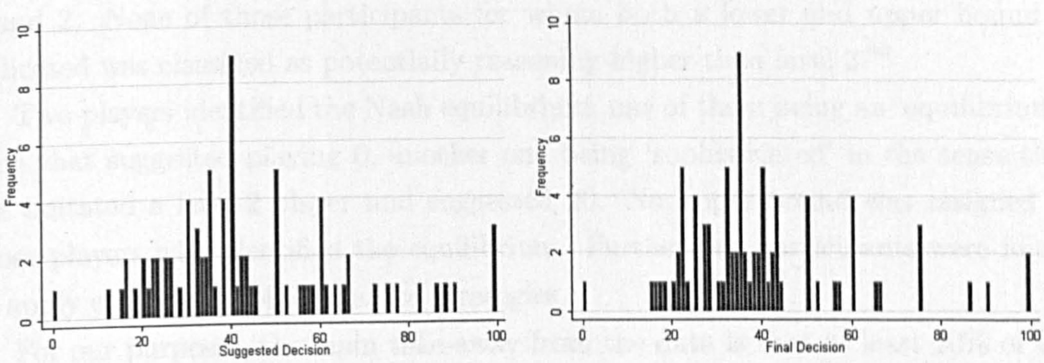
[93]The 6 subjects that did not write a message are dropped from this exposition.

[94]The spike at 40 might be unusual. The communication data reveals that this arises mainly as a result of two factors: some level-0 players chose 40 and some level-1 players chose 40 as they held the belief that the level-0 mean would equal 60. This insight shows how relaxing the dependence on action data allows for an analysis of the structural characteristics of reasoning.

TABLE 24: SUMMARY STATISTICS

	Mean	Std. Dev.	Median	Min.	Max.	N
PANEL A: Full Sample						
Suggested Decision	43.93	21.14	40	0	100	84
Final Decision	39.73	18.75	35	0	100	84
Team Action^a	40.02	18.98	35.5	16	100	84
PANEL B: Level-0 players						
Level-0 Action	62.35	22.39	60	16	100	17
PANEL C: Non-level-0 players with stated belief						
Level-0 Belief	55.26	12.33	50	40	100	36

Notes: [a] The team action is a random draw of the two final decisions.



(A) SUGGESTED DECISIONS (B) FINAL DECISIONS

FIGURE 12: INDIVIDUAL DECISIONS IN THE EXPERIMENT

– slightly higher action choices is reassuring in this respect.

The final decision of the participants has a lower mean and median than the suggested decision. This supports the intuition that the group-decision making process increases the level of reasoning and hence, in the 'beauty contest' game, leads to lower chosen numbers on average. Moreover the standard deviation of decisions across participants decreases after the exchange of the message and outliers become fewer, consistent with the idea that level-0 players who potentially chose the outliers, become fewer.[95]

3.5.2 Level-k Bounds

Table 25 presents the lower and upper bounds on the level of reasoning of individuals. For 70 participants both a lower and an upper bound was indicated. Eight participants have a non-classified upper bound. Another 6 participants did not make any statement and could therefore not be classified.

For 50 of the 84 participants the lower and the upper bounds coincide ($\approx 60\%$)

[95] An analysis of this process of persuasion is presented in Penczynski (2010b).

TABLE 25: LEVEL CLASSIFICATION RESULTS

		Level upper bounds					
		0	1	2	3	NA	Total
	0	17	11	1	0	6	35
Level	1		26	3	0	2	31
lower	2			6	5	0	11
bounds	3				1	0	1
	NA					6	6
	Total	17	37	10	6	14	84

and hence the classification fully determines their level of reasoning. These are 17 level-0, 26 level-1, 6 level-2 and 1 level-3 players, corresponding to the diagonal of Table 25. For further 20 players the classification restricts the level of reasoning to be one of two possibilities. Only for one participant we have an interval between 0 and 2. None of those participants for whom both a lower and upper bound is indicated was classified as potentially reasoning higher than level 3.[96]

Two players identified the Nash equilibrium, one of them being an 'equilibrium'-type that suggested playing 0, another one being 'sophisticated' in the sense that she imitated a level-2 player and suggested 20. No upper bound was assigned to those players who identified the equilibrium. Further two participants were found to apply elimination of dominated strategies.

For our purposes, the main take-away from the data is that at least 20% of the subjects are non-strategic reasoners, since 17 out of 84 participants are identified as level-0 reasoners. This is, however, only a lower bound on the fraction of level-0 reasoners, since for an additional 24 participants we cannot exclude the possibility that they are level-0 reasoners. When estimating the level-k model, we estimate over one third of the population to be level-0 reasoners (37%, see section 3.6.3).

3.5.3 Level-0 Action

We analyse the actions chosen by those with a lower and upper bound of 0 in order to provide an estimate of the distribution of level-0 play. Panel B of table 24 shows summary statistics of the suggested decisions of level-0 players and figure 13 presents the corresponding histogram. A one sample Kolmogorov-Smirnov test suggests that the distribution is significantly different from uniform (p-value= 0.038). Level-0 players choose on average 62 and their median choice is 60. The hypothesis that the mean of their choices is equal to 50 is rejected when tested against the alternative hypothesis of a higher mean (t-test, p-value= 0.019).

We find a higher frequency of level-0 actions around 50, 66 and 100. These can be seen as focal points in the spirit of Schelling (1960): 50 and 100 are focal due to the action space being integers between 0 and 100 and 66 may be focal due

[96]The data by subject can be obtained from the authors upon request.

to the multiplier in the game being 2/3. This adds to the evidence provided by Bacharach and Stahl (2000) and Crawford and Iriberri (2007a), who argue that salience considerations importantly shape the action distribution of non-strategic players.[97] However, the games they analyse display purposely framed actions. This is not true for the 'beauty contest' game. The fact that we still find an action distribution markedly different from uniform and centered around certain numbers, is consistent with an idea presented by Lewis (1969, p. 35) on the role of salience: he hypothesises that individuals "tend to pick the salient as a last resort, when they have no stronger ground for choice." Level-0 players can be understood as having no stronger reasons for choice than salience.

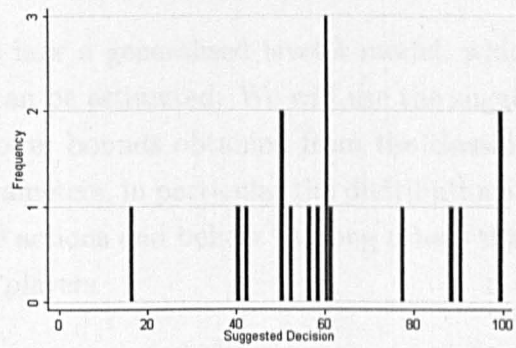


FIGURE 13: SUGGESTED DECISIONS OF LEVEL-0 PLAYERS

3.5.4 Level-0 Beliefs

The analysis of the written accounts of individual reasoning further allows us to analyse the players' beliefs about the average action of level-0 players. In 36 of the messages the players stated a non-derived belief about the average action of other players. Table 24 shows summary statistics of the stated level-0 beliefs. The mean belief is significantly higher than 50 (one-sided t-test, p-value= 0.007). Figure 14 presents the distribution of the level-0 beliefs.

More than 20 participants started reasoning with a level-0 belief of exactly 50. However, 11 players have a level-0 belief between 55 and 66. We interpret this as evidence that level-0 beliefs are indeed heterogeneous. This provides an explanation for why the actions of higher level, same-level players are heterogeneous.

Secondly, and to us surprisingly, the data suggests that level-0 beliefs are closely linked to the distribution of level-0 actions. In particular, they might pick up the same salience considerations which are found in the level-0 actions. Further, we fail to reject the hypothesis that the mean of the level-0-action and -belief distributions as shown in figures 13 and 14 are the same (two sample t-test with unequal variances, p-value: 0.235).

[97]Penczynski (2010a) uses the present paper's method to illuminate the level-0 distribution and belief in the context of the 'hide and seek' game.

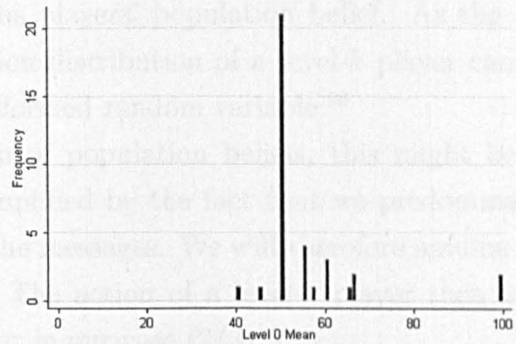


FIGURE 14: LEVEL-0 BELIEFS FROM COMMUNICATION TRANSCRIPTS

3.6 Estimation Of Structural Level-k Model

This section discusses how a generalised level-k model, which allows for heterogeneous level-0 beliefs, can be estimated. We will use the suggested decision together with the upper and lower bounds obtained from the classification to estimate the model's structural parameters, in particular the distribution of level-k types and the distributions of level-0 actions and beliefs. Among others this allows us to estimate the fraction of level-0 players.

3.6.1 An Estimable Model

The level-k model outlined in section 3.2 makes a probabilistic prediction about the observed actions. Let $f_j(x\,|\,\boldsymbol{\theta}_j)$ denote the probability mass function over the actions of a level j player. Then the unconditional probability mass function of the action of some player i can be written as

$$p(x_i\,;\,\psi) = \sum_{j=1}^{k} l_j f_j(x_i\,|\,\boldsymbol{\theta}_j) \tag{31}$$

where $\psi = (\boldsymbol{l}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., , \boldsymbol{\theta}_n)$, $l_j \geq 0$ for all j and $\sum_{j=1}^{k} l_j = 1$. This is a convex combination of component densities denoted in the statistics literature as 'finite mixture distribution'. The l_j's give the weight of each component distribution, and the $\boldsymbol{\theta}_j$'s are the parameters which characterize each component distribution. The distribution of all actions is hence a finite mixture of the action distributions of level-0, level-1, level-2 players and so on. We will outline in section 3.6.2 how to consistently estimate the parameter vector ψ. In the following we will first describe which form $f_j(x\,|\,\boldsymbol{\theta}_j)$ takes, how it depends on the level-0 action and belief distributions and how we parameterise these.

Action distributions in the level-k model The action distribution of a level-0 player, $f_0(x\,|\,\boldsymbol{\theta}_0)$, is simply the level-0 distribution $g^0(x\,|\,\boldsymbol{\theta}^0)$. The actions of higher level players, $f_j(x\,|\,\boldsymbol{\theta}_j)$ with $j \geq 1$, are derived from their level-0 belief, $g^b(x\,|\,\boldsymbol{\theta}^b)$, taking into account the players' population belief. As the level-0 belief is a random variable, the action distribution of a level-k player can be understood as the distribution of a transformed random variable.[98]

For a general form of population beliefs, this might be a complicated transformation. This is simplified by the fact that we predominantly found degenerate population beliefs in the messages. We will therefore assume degenerate population beliefs throughout.[99] The action of a level-k player then follows from his level-0 belief, say b^*, as $p^k b^*$ or in our case $(2/3)^k b^*$.

Parametric assumptions We parameterise $g^0(x \mid \boldsymbol{\theta}^0)$ as bounded normal distribution defined on the interval $[0, 100]$ and characterised by the mean and standard deviation $(\mu^0, \sigma^0) = \boldsymbol{\theta}^0$. This allows for a concentration of the level-0 actions as well as an approximate uniform distribution when σ^0 is large. We parameterise the distribution of the level-0 belief as well as bounded normal defined on the interval $[0, 100]$ and characterised by the mean and standard deviation $(\mu^b, \sigma^b) = \boldsymbol{\theta}^b$. As a special case this allows for the level-0 belief to be concentrated on the mean of the level-0 action distribution. To find the action distribution of level-k players that follows from this level-0 belief distribution, Lemma 1 in the appendix is useful. It states that a random variable which is distributed as bounded normal on $[0, 100]$ with parameters (μ, σ) will – when applying a multiplicative transformation using a factor a – be distributed as bounded normal with parameters $(a\mu, a\sigma)$ and support $[0, a100]$. Therefore we find the action distribution $f_k(x \mid \boldsymbol{\theta}_k)$, $k \geq 1$, as a bounded normal distribution with $\boldsymbol{\theta}_k = (2/3)^k \boldsymbol{\theta}^b$ and support $[0, (2/3)^k 100]$.[100]

We estimate the fraction of level reasoners for levels 0 to 3. No parametric structure is imposed on the distribution of level-reasoners, but we assume the highest level of reasoning observed to be 3. We do not find any communication that hints to a level of reasoning higher than that.

3.6.2 Estimator and Identification

Likelihood Function Given the probability mass function in equation 31 for the action of a player of unknown level, we can write the log-likelihood of the data as

$$L(\boldsymbol{x}; \boldsymbol{\psi}) = \sum_{i=1}^{n} \log p(x_i;\, \boldsymbol{\psi}) \tag{32}$$

[98]Note that a non-degenerate level-0 belief distribution will imply a non-degenerate action distribution for higher levels. This is in contrast to other models, where only if a player exactly matches the point prediction of the level-k model he would be classified as a higher level player.

[99]9 out of 12 players exhibited a degenerate population belief.

[100]Note that in equation 31 the component density f_j is indexed, allowing for the possibility that they are of different parametric families. In the case of the 'beauty contest' we can omit the subscript.

where $\boldsymbol{x}$ is the vector of actions observed. We use the information on the bounds by imposing $l_j = 0$ in $p(x_i; \boldsymbol{\psi})$ when the classification information is such that individual i is certainly not of level j.

Identification For the model to be identified, the mixture densities need to be linearly independent for all mixture probabilities $l_j \neq 0$. In the 'beauty contest' game with a parameterisation of the level-0 action and belief distribution as bounded normal this will necessarily be satisfied, irrespective of how many levels are estimated.[101]

Maximum Likelihood Estimator We estimate the parameter vector $\boldsymbol{\psi}$ with the maximum-likelihood estimator, denoted $\hat{\boldsymbol{\psi}}_{MLE}$. The log-likelihood function is thrice differentiable and the expectation of the third partial derivative is finite. We are unable to calculate the true information matrix, but we calculate an estimate of it and verify that it is positive definite. The MLE for this estimation problem is hence consistent, asymptotically normal and with asymptotic variance given by the inverse of the Fisher information matrix. We find the global maximiser of the log-likelihood function numerically, since the likelihood equations cannot be solved analytically. Details of this procedure are given in appendix C.2.

In order to ensure that the estimator is unbiased in a small sample, we ran Monte Carlo studies of sample size $N = 84$. For these we have generated data with the model given in equation 31, using $\hat{\boldsymbol{\psi}}_{MLE}$ for $\boldsymbol{\psi}$. The results give no reason to believe that our estimator is biased. Details are given in appendix C.4.

3.6.3 Estimation Results

Level-k distribution Table 26 shows the estimation results for the level-k distribution. We estimate 47% of the participants to be level-1 reasoners, and 15% to be level-2 reasoners. We estimate only 1% of the participants to be level-3 reasoners. This is similar to the classification results in section 3.5.2.[102] Crucially, we estimate that more than one third of the players (37%) are level-0 players.

The estimates in Table 26 on the relative frequencies of level-k reasoners, conditional on $k \geq 1$, resemble the relative fractions of level-1, level-2 and level-3 reasoners found in the literature for various games (see Camerer, Ho, and Chong (2004), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007a) etc.). This is reassuring and adds to the earlier evidence that the level-k model can have predictive power for the distribution of actions as a function of population and game characteristics.

[101]Unless in the special case where $\boldsymbol{\theta}^0 = (2/3)^k \boldsymbol{\theta}^b$ for some $k \geq 1$. For for any non-zero level-0 action mean, this will not be true if the mean of the level-0 belief is at the mean of the level-0 action.

[102]For computational reasons, the maximum level in the 'beauty contest' estimation is level-3.

TABLE 26: ESTIMATED LEVEL-k DISTRIBUTION

Parameter	l_0	l_1	l_2	l_3
Estimate	0.37	0.47	0.15	0.01
	(0.057)	(0.058)	(0.042)	(0.016)

Notes: The table presents the results from a maximum likelihood estimation of the structural model as outlined in section 3.6.1. This table only presents the results for the level-k distribution, but the level-0 action and belief distribution were estimated simultaneously. Those results are reported in table 27. Bootstrapped standard errors are given in brackets. These are obtained from 200 iterations of our estimation when sampling 84 observations from our data.

However, our estimate of the fraction of non-strategic reasoners is substantially higher than previously estimated. Nagel (1995) associates certain actions with level-0 play and estimates between 2% and 17% of the population to be level-0 reasoners. Camerer, Ho, and Chong (2004) parameterise the type distribution as Poisson and estimate the mean to be roughly 1.5, corresponding to roughly 22% non-strategic play. Note that our estimation procedure does not conflate level-0 play and errors of higher-level players, since we de-couple the level-0 action and belief distribution. Another way of interpreting the heterogeneity in the level-0 beliefs in the estimation is to think of it as errors of higher level players.[103]

Some of the level-k literature suggests that level-0 players only exist in the heads of other players. The evidence presented here sheds substantial doubts on this assumption. Of course, studies differ in the amount of testing that is done before the experiment, which might influence the capability of players to play strategically. In our study, we took care not to train or hint towards any strategic consideration. In our view, this should be the approach for studying one shot games.

Level-0 actions and beliefs Table 27 presents our estimates of the parameters characterising the distribution of actions of level-0 players and the beliefs of higher level players regarding their play.

TABLE 27: ESTIMATED LEVEL-0 ACTIONS AND BELIEFS

Parameter	μ^0	σ^0	μ^b	σ^b
Estimate	58.38	19.73	54.01	16.28
	(7.09)	(3.45)	(2.49)	(2.41)

Notes: The table presents the results from a maximum likelihood estimation of the structural model as outlined in section 3.6.1. This table only presents the results for the level-0 action and belief distribution, but the level-k distribution was estimated simultaneously. Those results are reported in table 26. Bootstrapped standard errors are given in brackets. These are obtained from 200 iterations of our estimation when sampling 84 observations from our data.

We estimate the mean of the level-0 action distribution to be 58.38 and the mean

[103] However, we want to emphasise, that we indeed find substantial heterogeneity in the stated level-0 beliefs in the classification procedure.

of the level-0 belief distribution to be 54.01. These are close to the estimates we obtained non-parametrically in Section 3.5. We estimate the level-0 action distribution to have a variance of 19.73 and the level-0 belief distribution to have a variance of 16.28. These are as well similar to the non-parametric estimates obtained before. Hence, we again find the belief distribution to be more concentrated than the action distribution, but – in contrast to the assumption made in the literature[104] – we estimate the belief distribution to have substantial variance.[105]

3.6.4 Quantifying the Classification Information

We can quantify the benefit of the additional information from our design by comparing the variance of our estimator with the variance of the equivalent estimator which does not use classification information.[106] Under suitable regularity conditions the maximum-likelihood estimator $\hat{\psi}_{MLE}$ will have an asymptotic distribution with variance matrix $\boldsymbol{I}^{-1}$, where $\boldsymbol{I}$ is the Fisher information matrix. The finite sample distribution then has an approximate variance matrix $\boldsymbol{I}^{-1}/n$. Let $\boldsymbol{I}_0^{-1}$ denote the information matrix of the estimator using unclassified data and $\boldsymbol{I}_c^{-1}$ the information matrix of our estimator which uses the partly classified data. Then $\boldsymbol{I}_0^{-1}/n$ and $\boldsymbol{I}_c^{-1}/n$ are the approximate finite variance matrices of these estimators, respectively. We can calculate how many more unclassified data points one would need to achieve the same efficiency as our design. As the variances are matrices, one needs to choose some real-valued summary statistic of the matrix to calculate an exact ratio. It is a common procedure in the literature on optimal experimental design to compare the traces of the variance matrices, also known as the 'total sum of variances'. As a bottom line, to estimate μ^0, σ^0, μ^b and σ^b with the same sum of variances, one would need 13 times as many choice observations as we have classified observations. Furthermore to estimate the l_j's with the same precision one would need 17 times as many observations. In our case, this would correspond to more than 1000 and 1400 observations respectively.[107]

3.7 Concluding Comments

In this study we present a novel experimental design which allows to investigate the anchoring element of the level-k model, the level-0 actions and beliefs. In a standard 'beauty contest' game, we find that one third of the participants are playing non-strategically. The level-0 beliefs are mostly 50, but significant upward deviations

[104] Recall that the literature assumes the level-0 action to be uniformly distributed and the level-0 belief to be degenerate at 50.

[105] When estimating the level-0 action distribution with a beta distribution that nests the uniform distribution, the estimated distribution is very similar in shape to the one found here.

[106] For references on the literature of optimal experimental design with mixture densities, check for example Hosmer and Dick (1977).

[107] Appendix C.5 gives details on the estimation and estimates of $\boldsymbol{I}_0^{-1}$ and $\boldsymbol{I}_c^{-1}$.

occur. On average, the beliefs are consistent with the non-uniform level-0 actions, which are concentrated around points 50, 66 and 100. These numbers suggest that one plausible determinant for level-0 actions is salience.

The novel features of our design are that the participants state their individual reasoning, (i) in close temporal proximity to the reasoning process itself, and (ii) are supplied an incentive to state their reasoning as fully and clearly as possible.

The written accounts provide an immediate insight into the individual's reasoning in addition to the participant's action in the game. This relaxes the need to make assumptions on aspects of the reasoning in order to interpret the player's action. The level of reasoning, for example, can be determined without particular assumptions on the level-0 belief or population belief distributions. In addition, it is possible to directly observe level-0 play or level-0 beliefs without any preconception of their distributions.

Obtaining an incentivised written account of individual reasoning also relaxes the need to design complex games with a view to drawing inference from actions alone. It allows us to learn about individual reasoning in games which are economically interesting even when actions alone are not informative. The design is generally applicable to one shot games and we believe can prove useful for other purposes in experimental economics.

C Appendix

C.1 Transformation of Bounded Normal

Lemma 1 *A random variable which is distributed as bounded normal on* $[0, 100]$ *with parameters* (μ, σ) *will still be distributed as bounded normal when applying a multiplicative transformation with a factor* a*, with parameters* $(a\mu, a\sigma)$ *and support* $[0, a100]$.

Proof. For a monotonic, differentiable transformation of a random variable B, say $X = g(B)$, the distribution of X is found as $f_X(x) = f_B(g^{-1}(x)) \cdot |g^{-1'}(x)|$ where the support needs to be suitably changed. In our case $g(b) = p^k b$, $g^{-1}(x) = x/p^k$ and $g^{-1'}(x) = 1/p^k$

$$f(x; \bullet) = \frac{\frac{1}{p^k \sigma^b} \phi(x/p^k; \boldsymbol{\theta}^b)}{\Phi(100; \boldsymbol{\theta}^b) - \Phi(0; \boldsymbol{\theta}^b)} = \frac{\frac{1}{p^k \sigma^b} \phi(y; p^k \boldsymbol{\theta}^b)}{\Phi(p^k 100; p^k \boldsymbol{\theta}^b) - \Phi(0; p^k \boldsymbol{\theta}^b)}$$

where the equality of the nominators follows by straight-forward manipulation of $\phi()$ and the denominator follows from

$$\Phi(T; \boldsymbol{\theta}^b) = \int_{-\infty}^{T} \frac{1}{\sigma^b} \cdot \phi\left(\frac{x - \mu^b}{\sigma^b}\right) \mathrm{dz} = \int_{-\infty}^{p^k T} \frac{1}{p^k \sigma^b} \cdot \phi\left(\frac{w - p^k \mu^b}{p^k \sigma^b}\right) \mathrm{dw} = \Phi\left(p^k T; p^k \boldsymbol{\theta}^b\right)$$

after a change of variable $w = p^k z$. ■

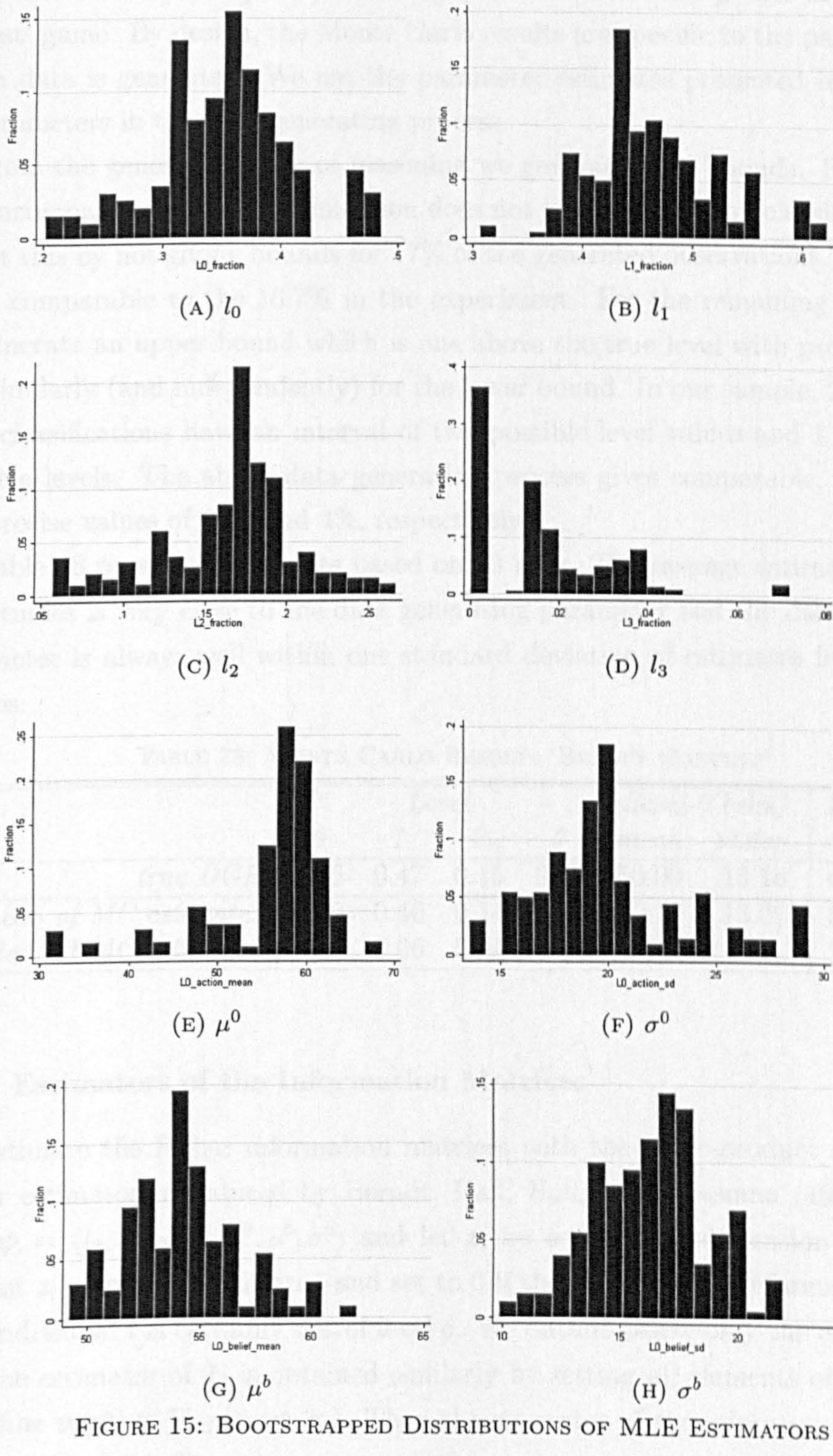
C.2 Details of the Numerical Estimation

In order to find the maximum likelihood estimates, we numerically maximise the log-likelihood. We start out searching over a grid covering the the full range of parameter values.[108] Once the vector of parameter values is found which gives the highest likelihood, a new, finer grid is defined around this set. The new grid includes the parameter values that were neighbors of the maximising set in the previous, coarser grid. The calculation is iterated until the mesh size is sufficiently small.

C.3 Bootstrapping

The standard errors for the maximum likelihood estimates have been bootstrapped. Estimating 200 samples that consist of 84 draws with replacement from the original dataset, we obtain a measure of the estimates's variance. Figure 15 shows the histograms of the 200 estimates for the 8 estimated parameters.

[108] All sets of possible parameter values are bounded except for the standard deviation of the bounded normal distribution. Using a beta distribution allows us to cover distribution shapes that are equivalent to standard deviations up to infinity in the bounded normal case.



(A) l_0 (B) l_1

(C) l_2 (D) l_3

(E) μ^0 (F) σ^0

(G) μ^b (H) σ^b

FIGURE 15: BOOTSTRAPPED DISTRIBUTIONS OF MLE ESTIMATORS

C.4 Monte Carlo Studies

We use Monte Carlo studies in order to analyse the properties of our maximum likelihood estimators. We use the estimable model outlined in section 3.6.1 to generate data for 84 participants, reflecting the number of data points in the 'beauty contest' game. By design, the Monte Carlo results are specific to the parameters for which data is generated. We use the parameter estimates presented in section 3.6 as parameters in the data generating process.

From the generated levels of reasoning we generate level bounds. Firstly, there are participants whose communication does not lend itself to any classification. We reflect this by not giving bounds for 17% of the generated observations, making this value comparable to the 16.7% in the experiment. For the remaining data points we generate an upper bound which is one above the true level with probability 0.2 and similarly (and independently) for the lower bound. In our sample, 27.1% of the level classifications have an interval of two possible level values and 1.4% of three possible levels. The above data generating process gives comparable, but slightly less precise values of 32% and 4%, respectively.

Table 28 presents the results based on 50 runs. The average estimate from the MC studies is very close to the data generating parameter and the data generating parameter is always well within one standard deviation of estimates from the MC studies.

TABLE 28: MONTE CARLO RESULTS 'BEAUTY CONTEST'.

	Level				*Level-0 belief*		*Level-0 action*	
	0	*1*	*2*	*3*	*mean*	*st.dev.*	*mean*	*st.dev.*
true DGP	0.36	0.47	0.15	0.02	50.00	13.16	60.00	21.00
mean of MC estimates	0.39	0.46	0.14	0.01	50.13	13.07	56.45	21.17
st. dev. of MC estimates	0.05	0.06	0.04	0.01	1.95	1.68	5.45	4.80

C.5 Estimators of the Information Matrices

We estimate the Fisher information matrices with the outer-product of the score vector estimator introduced by Berndt, Hall, Hall, and Hausman (1974). Recall that $\boldsymbol{\psi} = (l_0, l_1, l_2, \mu^0, \sigma^0, \mu^b, \sigma^b)$ and let $\boldsymbol{z}_i$ be a vector of dimension 1×4 with element z_{ij} generally equal to 1 and set to 0 if the classification information is such that individual i is certainly not of level j. We outline below only the estimation of $\boldsymbol{I}_c$. The estimator of $\boldsymbol{I}_0$ is obtained similarly by setting all elements of $\boldsymbol{z}_i$ equal to 1. Define $\boldsymbol{v}_i(\boldsymbol{\psi}) \equiv \nabla_{\boldsymbol{\psi}} p(x_i; \boldsymbol{\psi}, \boldsymbol{z}_i)$. Then the estimator of the information matrix $\boldsymbol{I}_c$ is $\sum_i \boldsymbol{v}_i(\boldsymbol{\psi})\boldsymbol{v}_i(\boldsymbol{\psi})'$. The score vector $\boldsymbol{v}_i(\boldsymbol{\psi})$ has elements

$$\begin{aligned}
\frac{\partial p_i}{\partial l_0} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i0}f(x_i;\boldsymbol{\theta}^0) - z_{i3}f(x_i;p^3\boldsymbol{\theta}^b)\right]\\
\frac{\partial p_i}{\partial l_1} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i1}f(x_i;p^1\boldsymbol{\theta}^b) - z_{i3}f(x_i;p^3\boldsymbol{\theta}^b)\right]\\
\frac{\partial p_i}{\partial l_2} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i2}f(x_i;p^2\boldsymbol{\theta}^b) - z_{i3}f(x_i;p^3\boldsymbol{\theta}^b)\right]\\
\frac{\partial p_i}{\partial \mu^0} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i0}l_0\frac{\partial f(x_i;\boldsymbol{\theta}^0)}{\partial \mu^0}\right]\\
\frac{\partial p_i}{\partial \sigma^0} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i0}l_0\frac{\partial f(x_i;\boldsymbol{\theta}^0)}{\partial \sigma^0}\right]\\
\frac{\partial p_i}{\partial \mu^b} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i1}l_1\frac{\partial f(x_i;p^1\boldsymbol{\theta}^b)}{\partial \mu^b} + z_{i2}l_2\frac{\partial f(x_i;p^2\boldsymbol{\theta}^b)}{\partial \mu^b} + z_{i3}(1-l_0-l_1-l_2)\frac{\partial f(x_i;p^3\boldsymbol{\theta}^b)}{\partial \mu^b}\right]\\
\frac{\partial p_i}{\partial \sigma^b} &= \frac{1}{p(x_i;\boldsymbol{\psi},\boldsymbol{z}_i)}\left[z_{i1}l_1\frac{\partial f(x_i;p^1\boldsymbol{\theta}^b)}{\partial \sigma^b} + z_{i2}l_2\frac{\partial f(x_i;p^2\boldsymbol{\theta}^b)}{\partial \sigma^b} + z_{i3}(1-l_0-l_1-l_2)\frac{\partial f(x_i;p^3\boldsymbol{\theta}^b)}{\partial \sigma^b}\right].
\end{aligned}$$

We find

$$\begin{aligned}
\frac{\partial f(x;a\boldsymbol{\theta})}{\partial \mu} &= \frac{\frac{1}{(a\sigma)}\phi(x,a\boldsymbol{\theta})\left[\frac{x-a\mu}{a\sigma^2}[\Phi(a100;a\boldsymbol{\theta})-\Phi(0;a\boldsymbol{\theta})] - \frac{1}{\sigma}[\phi(0;a\boldsymbol{\theta})-\phi(a100,a\boldsymbol{\theta})]\right]}{[\Phi(a100;a\boldsymbol{\theta})-\Phi(0;a\boldsymbol{\theta})]^2}\\
&= f(x;a\theta)\frac{1}{\sigma}\left[\frac{x-a\mu}{a\sigma} + a\sigma(f(a100;a\theta) - f(0;a\theta))\right]
\end{aligned}$$

where the second step follows by the definition of $f(x;a\theta)$. Secondly we find

$$\frac{\partial f(x;a\boldsymbol{\theta})}{\partial \sigma} = \frac{\frac{\partial \frac{1}{a\sigma}\phi(x,a\theta)}{\partial\sigma}}{[\Phi(a100;a\theta)-\Phi(0;a\theta)]} - \frac{\frac{1}{a\sigma}\phi(x,a\theta)\frac{\partial[\Xi_0]}{\partial\sigma}}{[\Phi(a100;a\theta)-\Phi(0;a\theta)]^2}$$

where $\Xi_0 = \Phi(a100;a\theta) - \Phi(0;a\theta)$. We can calculate

$$\frac{\partial \frac{1}{a\sigma}\phi(x,a\theta)}{\partial\sigma} = \frac{1}{\sigma}\frac{1}{a\sigma}\phi(x;a\theta)\left[\frac{(x-a\mu)^2}{a^2\sigma^2} - 1\right]$$

and

$$\begin{aligned}
\frac{\partial[\Phi(a100;a\theta)-\Phi(0;a\theta)]}{\partial\sigma} &= \frac{\partial\int_0^{a100}\frac{1}{a\sigma}\phi(x;a\theta)\mathrm{dx}}{\partial\sigma}\\
&= \int_0^{a100}\frac{\partial\frac{1}{a\sigma}\phi(x;a\theta)}{\partial\sigma}\mathrm{dx}\\
&= \int_0^{a100}\frac{1}{\sigma}\frac{1}{a\sigma}\phi(x;a\theta)\left[\frac{(x-a\mu)^2}{a^2\sigma^2} - 1\right]\mathrm{dx}\\
&= \frac{1}{\sigma}\int_0^{a100}\frac{(x-a\mu)^2}{(a\sigma)^2}\frac{1}{a\sigma}\phi(x;a\theta)\mathrm{dx} - \frac{1}{\sigma}[\Phi(a100;a\theta) - \Phi(0;a\theta)]
\end{aligned}$$

Divide by $\Phi(a100;a\theta) - \Phi(0;a\theta)$ to find

$$\begin{aligned}
&\frac{1}{\sigma^3 a^2}\int_0^{a100}(x^2 - 2a\mu x + a^2\mu^2) f(x;a\theta)\mathrm{dx} - \frac{1}{\sigma} \\
&= \frac{1}{\sigma^3 a^2}\underbrace{\int_0^{a100} x^2 f(x;a\theta)\mathrm{dx}}_{\Xi_2} - \frac{2a\mu}{\sigma^3 a^2}\underbrace{\int_0^{a100} x f(x;a\theta)\mathrm{dx}}_{\Xi_1} + \frac{a^2\mu^2}{\sigma^3 a^2}\underbrace{\int_0^{a100} f(x;a\theta)\mathrm{dx}}_{=1} - \frac{1}{\sigma}
\end{aligned}$$

where Ξ_1 and Ξ_2 are the first and second central moments of the bounded normal. From the m.g.f. of the bounded normal they are found as

$$\begin{aligned}
\Xi_1 &= a\mu - a\sigma\frac{\phi(a100;a\theta) - \phi(0;a\theta)}{\Phi(a100;a\theta) - \Phi(0;a\theta)} \\
\Xi_2 &= a^2\mu^2 + a^2\sigma^2 + a^2\sigma^2\frac{\phi'(a100;a\theta) - \phi'(0;a\theta)}{\Phi(a100;a\theta) - \Phi(0;a\theta)} - 2a^2\mu\sigma\frac{\phi(a100;a\theta) - \phi(0;a\theta)}{\Phi(a100;a\theta) - \Phi(0,a\theta)}
\end{aligned}$$

Note that $\phi'(a100;a\theta) = -\frac{a100-a\mu}{a\sigma}\phi(a100;a\theta)$ and $\phi'(0;a\theta) = -\frac{0-a\mu}{a\sigma}\phi(0;a\theta)$. Collecting terms we find

$$\frac{\partial\left[\Phi(a100;a\theta) - \Phi(0;a\theta)\right]/\partial\sigma}{\Phi(a100;a\theta) - \Phi(0;a\theta)} = -\frac{1}{\sigma}\frac{\frac{a100-a\mu}{a\sigma}\phi(a100;a\theta) - \frac{0-a\mu}{a\sigma}\phi(0;a\theta)}{\Phi(a100;a\theta) - \Phi(0;a\theta)}$$

Collecting terms and substituting in $f(x;a\theta)$ we can write

$$\frac{\partial f(x;a\theta)}{\partial\sigma} = f(x;a\theta)\frac{1}{\sigma}\left[\left(\frac{x-a\mu}{a\sigma}\right)^2 - 1 + (a100 - a\mu)\cdot f(a100,a\theta) - (0 - a\mu)\cdot f(0,a\theta)\right]$$

C.6 Estimates of Information Matrices

Our estimate of the information matrix when not using information on the upper and lower bounds of reasoning is

$$\hat{I}_0 = \begin{bmatrix}
0.428 & -0.595 & 0.049 & -29.234 & 17.102 & 6.466 & -12.630 \\
-0.595 & 0.896 & -0.127 & 41.296 & -25.289 & -11.914 & 17.787 \\
0.049 & -0.127 & 0.064 & -3.800 & 3.095 & 3.129 & -1.504 \\
-29.234 & 41.296 & -3.800 & 2041.117 & -1184.958 & -469.305 & 869.482 \\
17.102 & -25.289 & 3.095 & -1184.958 & 751.216 & 334.845 & -515.823 \\
6.466 & -11.914 & 3.129 & -469.305 & 334.845 & 241.982 & -200.424 \\
-12.630 & 17.787 & -1.504 & 869.482 & -515.823 & -200.424 & 382.519
\end{bmatrix}$$

and our estimate of the information matrix when using this information is

$$\hat{I}_c = \begin{bmatrix} 0.018 & -0.007 & -0.004 & -0.684 & 0.341 & -0.309 & -0.064 \\ -0.007 & 0.021 & -0.010 & 0.417 & -0.148 & -0.021 & 0.096 \\ -0.004 & -0.010 & 0.042 & 0.162 & -0.136 & 0.191 & 0.030 \\ -0.684 & 0.417 & 0.162 & 116.136 & -17.443 & 15.078 & 3.438 \\ 0.341 & -0.148 & -0.136 & -17.443 & 83.307 & 2.554 & 3.215 \\ -0.309 & -0.021 & 0.191 & 15.078 & 2.554 & 40.129 & 9.249 \\ -0.064 & 0.096 & 0.030 & 3.438 & 3.215 & 9.249 & 8.512 \end{bmatrix}.$$

C.7 Classification Agreement

Table 29: CLASSIFICATION AGREEMENT

Subject	*Level-0*	*Level-1*	*Level-2*	*Level-3*	*Coinciding*
1	8				8
2			8		8
3	8				8
4	8				8
5	8				8
6		8			8
7	8				8
8	8				8
9	8				8
10		8			8
11	8				8
12		8			8
13		8			8
14	8				8
15		8			8
16	8				8
17	8				8
18	8				8
19			8		8
20	8				8
21	8				8
22	8				8
23	8				8
24		8			8

CONTINUED ON NEXT PAGE

Table 29 – CONTINUED FROM PREVIOUS PAGE

Subject	*Level-0*	*Level-1*	*Level-2*	*Level-3*	*Coinciding*
25		8			8
26	8				8
27	8				8
28	8				8
29	8				8
30		8			8
31	8				8
32	8				8
33		8			8
34				8	8
35	8				8
36		8			8
37		8			8
38		8			8
39		8			8
40	8				8
41	1	7			7
42			7	1	7
43	1	7			7
44	7	1			7
45			7	1	7
46		7	1		7
47	7	1			7
48	7	1			7
49	7	1			7
50	7	1			7
51		1	7		7
52	1	7			7
53	1	7			7
54	7	1			7
55	1	7			7
56	6	2			6
57	2	6			6
58	6	2			6
59	6	2			6
60	6	2			6

CONTINUED ON NEXT PAGE

Table 29 – CONTINUED FROM PREVIOUS PAGE

Subject	*Level-0*	*Level-1*	*Level-2*	*Level-3*	*Coinciding*
61		6	2		6
62	6	2			6
63	6	2			6
64	6	2			6
65		2	6		6
66		6	2		6
67	2	6			6
68	6	1	1		6
69		1	6	1	6
70	1	6	1		6
71	3	5			5
72	1	5	2		5
73		1	5	2	5
74	4	4			4
75		4	4		4
76	4	4			4
77	4	2	2		4
78	1	2	4	1	4

Notes: The table presents the number of classifiers (out of 8 classifiers) who picked the indicated lower bound, by subject. The subjects are ordered by the maximum number of coinciding choices.

References

AGRANOV, M., A. CAPLIN, AND C. TERGIMAN (2010): "The Process of Choice in Guessing Games," working paper, California Institute of Technology.

AKBULUT-YUKSEL, M. (2009): "Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children," working paper, University of Houston.

ALESINA, A., A. DEVLEESCHAUWER, W. EASTERLY, S. KURLAT, AND R. WACZIARG (2003): "Fractionalization," *Journal of Economic Growth*, 8.

ALESINA, A., AND E. L. FERRARA (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43(3), 762–800.

ALESINA, A., AND N. FUCHS-SCHÜNDELN (2007): "Good-Bye Lenin (or Not?): The Effect of Communism on People's Preferences," *The American Economic Review*, 97(4), 1507 - 1528.

AMBRUS, A., M. MOBIUS, AND A. SZEIDL (2010): "Consumption Risk-sharing in Social Networks," *NBER Working Paper Series 15719.*

ANDERSON, T. W., AND C. HSIAO (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

BACHARACH, M., AND D. O. STAHL (2000): "Variable-Frame Level-n Theory," *Games and Economic Behavior*, 32(2), 220–246.

BEAMAN, L. A. (2008): "Social networks and the dynamics of labor market outcomes: Evidence from refugees," working paper, Northwestern University.

BERNDT, E. K., B. H. HALL, R. E. HALL, AND J. A. HAUSMAN (1974): "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3/4, 653–665.

BESLEY, T., AND S. COATE (1995): "Group lending, repayment incentives and social collateral," *Journal of Development Economics*, pp. 1–18.

BETHLEHEM, S. (1982): *DDR-Flucht, Gastarbeiter-Zuwanderung.* Klett-Cotta, Stuttgart.

BOSCH-DOMÈNECH, A., J. G. MONTALVO, R. NAGEL, AND A. SATORRA (2002): "One, Two, (Three), Infinity, . . . : Newspaper and Lab Beauty-Contest Experiments," *The American Economic Review*, 92(5), 1687–1701.

——— (2004): "Finite Mixture Analysis of Beauty-Contest Data from Multiple Samples," Economics Working Papers 737, Department of Economics and Business, Universitat Pompeu Fabra.

BOSSERT, W., C. D'AMBROSIO, AND E. L. FERRARA (2010): "A Generalized Index of Fractionalization," *Economica*, pp. 1–29.

BRAKMAN, S., H. GARRETSEN, AND M. SCHRAMM (2004): "The Strategic Bombing of German Cities During WWII and its Impact on City Growth," *Journal of Economic Geography*, 4, 201–18.

BURSZTYN, L., AND D. CANTONI (2009): "Clueless? The impact of television on consumption behavior," working paper, Harvard University.

CAMERER, C. (2003): *Behavioral Game Theory*. Princeton University Press.

CAMERER, C., E. J. JOHNSON, S. SEN, AND T. RYMON (1993): "Cognition and framing in sequential bargaining for gains and losses," in *Frontiers of Game Theory*, ed. by A. K. Ken Binmore, and P. Tani, pp. 27–48. MIT Press.

CAMERER, C. F., T.-H. HO, AND J.-K. CHONG (2004): "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119(3), 861–898.

CENTRE DE RECHERCHES ET DE DOCUMENTATION SUR LA CONSOMMATION (1958): *Consommation*. Paris.

CHISWICK, B. R. (1998): "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel," *Journal of Population Economics*, 11(2), 253–271.

CHISWICK, B. R., AND P. W. MILLER (1995): "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13(2), 246–88.

——— (1999): "Language Skills and Earnings among Legalized Aliens," *Journal of Population Economics - Special Issue on Illegal Migration*, 12(1), 63–89.

CHISWICK, B. R., H. A. PATRINOS, AND M. E. HURST (2000): "Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia," *Economic Development and Cultural Change*, 48(2), 349–367.

COHEN, L., A. FRAZZINI, AND C. MALLOY (2008): "The small world of investing: Board connections and mutual fund returns," *Journal of Political Economy*, 116, 951–979.

CONLEY, T. G., AND C. R. UDRY (2009): "Learning about a new technology: Pineapple in Ghana," *forthcoming American Economic Review*.

COOPER, D. J., AND J. H. KAGEL (2005): "Are Two Heads Better than One? Team versus Individual Play in Signaling Games," *The American Economic Review*, 95(3), 477–509.

COSTA-GOMES, M. A., AND V. P. CRAWFORD (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *The American Economic Review*, 96(5), 1737–1768.

COSTA-GOMES, M. A., V. P. CRAWFORD, AND B. BROSETA (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69(5), 1193–1235.

COULMAS, F. (1992): *Language and Economics*. Blackwell Publishers, Oxford.

CRAWFORD, V. P., M. COSTA-GOMES, AND N. IRIBERRI (2010): "Strategic Thinking," working paper, University of Oxford.

CRAWFORD, V. P., U. GNEEZY, AND Y. ROTTENSTREICH (2008): "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yields Large Coordination Failures," *The American Economic Review*, 98(4), 1443–1458.

CRAWFORD, V. P., AND N. IRIBERRI (2007a): "Fatal Attraction: Focality, Naïveté and Sophistication in Experimental 'Hide and Seek' Games," *The American Economic Review*, 97(5), 1731–1750.

——— (2007b): "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75(6), 1721–1770.

DAVIS, D. R., AND D. E. WEINSTEIN (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *The American Economic Review*, 92(5), 1269–1289.

DUSTMANN, C., AND F. FABBRI (2003): "Language proficiency and labour market performance of immigrants in the UK," *Economic Journal*, 113(489), 695–717.

EASTERLY, W., AND R. LEVINE (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4), 1203–1250.

ERICSSON, K. (2002): "Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis," *Applied Cognitive Psychology*, 16(8), 981–987.

ESTEBAN, J., AND D. RAY (1994): "On the Measurement of Polarization," *Econometrica*, 62(4), 819–51.

FISCHBACHER, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

FRANZEN, K. E. (2001): *Die Vertriebenen.* Propylaen, Berlin.

FRIEDRICH, J. (2002): *Der Brand.* Propylaen, Berlin.

FUCHS-SCHÜNDELN, N., AND M. SCHÜNDELN (2005): "Precautionary Savings and Self-Selection: Evidence from the German Reunification Experiment," *Quarterly Journal of Economics*, 120(3), 1085–1120.

——— (2009): "Who stays, who goes, who returns?," *Economics of Transition*, 17(4), 703–738.

GODOY, R., V. REYES-GARCIA, C. SEYFRIED, T. HUANCA, W. R. LEONARD, T. MCDADE, S. TANNER, AND V. VADEZ (2007): "Language skills and earnings: Evidence from a pre-industrial economy in the Bolivian Amazon," *Economics of Education Review*, 26(3), 349–360.

GRANOVETTER, M. (1973): "The strength of weak ties," *American Journal of Sociology*, 78, 1360–1380.

HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669 738.

HOCHBERG, Y., A. LJUNGQVIST, AND Y. LU (2007): "Whom you know matters: Venture capital networks and investment performance," *The Journal of Finance*, 62, 251–301.

HOLBIK, K., AND H. MYERS (1964): *Postwar Trade in Divided Germany.* Johns Hopkins Press, Baltimore.

HOSMER, D. W., AND N. P. DICK (1977): "Information and mixtures of two normal distributions," *Journal of Statistical Computation and Simulation*, 6(2), 137–148.

HUNT, J. (2006): "Staunchig emigration from East Germany: Age and the determinants of migration," *Journal of the European Economic Associaton*, 4, 1014 1037.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES (1949): *Résultats Statistique Du Recensement Général de la Population 1946.* Imprimerie Nationale, Paris.

——— (1968): *Recensement de 1968.* Direction des Journaux Officiels, Paris.

——— (1999): *Edute de l'Histoire Familiale en 1999.* Institut National de la Statistique et des Études Économiques, Paris.

KRANTON, R. E., AND D. F. MINEHART (2001): "A Theory of Buyer-Seller Networks," *The American Economic Review*, 91(3), 485–508.

KUHNEN, C. M. (2009): "Business networks, corporate governance and contracting in the mutual fund industry," *forthcoming The Journal of Finance.*

KUROWSKI, F. (1977): *Der Luftkrieg ueber Deutschland.* Econ Verlag, Duesseldorf.

LANGE, T., AND G. PUGH (1998): *The Economics of German Reunification.* Edward Elgar, Cheltenham UK and Northampton MA.

LASCHEVER, R. A. (2007): "The doughboys network: Social interactions and the employment of World War I veterans," working paper, University of Illinois at Urbana-Champpaign.

LAZEAR, E. (1999): "Culture and Language," *Journal of Political Economy*, 107(6), 95–126.

LEWIS, D. K. (1969): *Convention: A Philosophical Study.* Harvard University Press.

MCKELVEY, R. D., AND T. R. PALFREY (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10(1), 6–38.

MIGUEL, E., AND G. ROLAND (2011): "The long-run impact of bombing Vietnam," *Journal of Development Economics*, 96(1), 1 – 15.

MONTALVO, J. G., AND M. REYNAL-QUEROL (2005a): "Ethnic diversity and economic development," *Journal of Development Economics*, 76, 293–323.

——— (2005b): "Ethnic Polarization, Potential Conflict, and Civil Wars," *The American Economic Review*, 95(3), 796–816.

MUNSHI, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market," *Quarterly Journal of Economics*, 118(2), 549–599.

NAGEL, R. (1994): "Reasoning and Learning in Guessing Games and Ultimatum Games With Incomplete Information: An Experimental Investigation," Ph.D. thesis, Universität Bonn.

——— (1995): "Unraveling in Guessing Games: An Experimental Study," *The American Economic Review*, 85(5), 1313–1326.

PENCZYNSKI, S. P. (2010a): "Focality, Tasks and the Beginning of Level-k Reasoning," working paper, Cornell University,.

——— (2010b): "Persuasion and Strategic Thinking," working paper, Cornell University.

RAUCH, J. E. (1999): "Networks versus markets in international trade," *Journal of International Economics*, 48(1), 7–35.

RAUCH, J. E., AND A. CASELLA (2003): "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties," *The Economic Journal*, 113(484), pp. 21–42.

REDDING, S. J., AND D. M. STURM (2008): "The Cost of Remoteness: Evidence from German Division and Reunifiaction," *The American Economic Review*, 98(5), 1766–97.

RUBINSTEIN, A. (2007): "Instinctive and Cognitive Reasoning: A Study of Response Times," *Economic Journal*, 117, 1243–1259.

RUBINSTEIN, A., A. TVERSKY, AND D. HELLER (1996): "Naive Strategies in Zero-Sum Games," in *Understanding Strategic Interaction–Essays in Honor of Reinhard Selten*, ed. by W. Albers, W. Güth, P. Hammerstein, B. Molduvano, and E. van Damme, pp. 394–402. Springer Verlag.

RYDVAL, O., A. ORTMANN, AND M. OSTATNICKY (2009): "Three very simple games and what it takes to solve them," *Journal of Economic Behavior & Organization*, 72(1), 589–601.

SCHELLING, T. C. (1960): *The Strategy of Conflict*. Harvard University Press.

SCHULZE, I. (2005): *Neue Leben*. Berlin Verlag, Berlin.

SHARP, T. (1975): *The Wartime Alliance and the Zonal Division of Germany*. Clarendon, Oxford.

SINN, G., AND H.-W. SINN (1992): *Jumpstart*. MIT Press, Cambridge MA.

SOUTHERN, D. B. (1993): "Restitution or Compensation: The Land Question in East Germany," *The International and Comparative Law Quarterly*, 42(3), 690 697.

STAHL, D. O., AND P. W. WILSON (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10(1), 218 254.

STATISTIQUE GÉNÉRALE DE LA FRANCE (1931a): *Annuaire Statistique 1930*. Imprimerie Nationale, Paris.

——— (1931b): *Résultats Statistique Du Recensement Général de la Population 1926*. Imprimerie Nationale, Paris.

——— (1951): *Annuaire Statistique 1950*. Imprimerie Nationale, Paris.

STIGLITZ, J. E. (1990): "Peer Monitoring and Credit Markets," *The World Bank Economic Review*, 4(3), 351–366.

U.S. GOVT. PRINT OFFICE (1945): *The United States Strategic Bombing Survey: the Effects of Strategic Bombing on the German War Economy*. Washington DC.

VARIAN, H. R. (1990): "Monitoring agents with other agents," *Journal of Institutional and Theoretical Economics (Zeitschrift fur die gesamte Staatswissenschaft)*, pp. 153–174.

WEERDT, J. D., AND S. DERCON (2006): "Risk-sharing networks and insurance against illness," *Journal of Development Economics*, 81(2), 337–356.