Those Who Stayed: Selection and Cultural Change in the Age of Mass Migration[*]

Anne Sofie Beck Knudsen[†]

September 15, 2021

Abstract

This paper studies the cultural determinants and consequences of mass emigration from Scandinavia to North America in the 19th century. I test the hypothesis that people with collectivist traits tended to stay rather than emigrate because they faced higher costs of leaving familiar social networks behind. Exploiting near-complete data on 1.5 million emigrants and 10 million stayers in Denmark, Norway, and Sweden, I find that children who grew up in households that practiced stronger collectivist norms were less likely to emigrate later in life. I proceed to document that this type of selective emigration generated lasting cultural change in migrant-sending locations. Locations that experienced larger outflows of particularly selected individuals are thus more collectivist today. The implications of these findings are potentially wide-reaching as collectivism and its counterpart, individualism, have been linked to central societal processes such as cooperation, the diffusion of ideas, and innovation.

Keywords: Migration, culture, identity, selection, Age of Mass Migration

JEL classification: F22, Z10, O15, R23, N33

[†]Department of Economics, University of Copenhagen, email: asbk@econ.ku.dk

1 Introduction

International migrants make up 3.5 percent of the global population, and the number is rising (IOM, 2020). Understanding the economic and political consequences of migration has been at the center of growing scholarly interest.[1] One dimension that has received less attention is the degree to which migration influences the evolution of cultures, particularly in the societies from which migrants leave.[2]

This paper studies whether culture plays a role in the individual decision to migrate, and how this type of selection can generate profound cultural change in the communities of origin. The motivation is a prominent hypothesis from social-psychology, which posits that people with collectivist cultural traits are less prone to migrate than their individualistic counterparts (Kitayama et al., 2006). Being more reliant on familiar networks for identity and support, collectivists face a higher cost of migration as it often involves leaving such networks behind. If culture is transmitted across generations, this type of selection in migration is expected to generate lasting cultural change: When individualists leave, culture at the origin becomes more collectivist on average. This may have wide societal implications as the individualism-collectivism cultural dimension has been linked to economic development through its impact on the diffusion of ideas, innovation, and scope of cooperation.[3,4]

I test the hypothesis and its implications in the setting of Scandinavia in one of the largest migration events in modern history: the Age of Mass Migration between 1850 and 1920. Combining both novel and existing data sources from Denmark, Norway, and Sweden, I test each element of the proposed relationship between migration and collectivism.

Several circumstances make the Age of Mass Migration an ideal case for the objective of this paper. In this period, more than 40 million people left Europe to settle in New World countries such as the United States. Emigration was spurred by a drop in transport costs associated with the shift from sail to steamship technology and the end of the American Civil War in 1865. Moreover, global immigration policies

---

[1]Recent studies on the long-run effects of migration include Hornung (2014), Droller (2018), Burchadi et al. (2019), Sequira et al. (2020), and Tabellini and Giuliano (2020).

[2]Research on this topic has generally focused on the places of migrants' arrival and short-run cultural dynamics. Studies on persistent cultural effects in places of arrival include Fernández and Fogli (2009), Grosjean (2014), and Fouka et al. (2021).

[3]See, among others, Greif (1994), Gorodnichenko and Roland (2011a, 2011b), Enke (2019), Buggle (2020), Henrich (2020), and Fogli and Veldkamp (2021).

[4]Moreover, if collectivism is a barrier to geographical mobility, this may impact the efficient allocation of labor in the economy. I discuss this further in connection with the related literature.

were particularly unrestrictive. Sweden, Norway, and Denmark experienced some of the highest emigration rates in Europe in this period, which involved the departure of approximately 25 percent of their populations. The historical dimension further enables the study of long-run cultural implications from 1850 to today.

For the empirical analysis, I rely on two sources of individual-level data. First, a collection of historical censuses that contain demographic and socioeconomic information on everyone who lived in Scandinavia at the time. Second, I compile a novel database on 1.5 million Scandinavians who emigrated in the Age of Mass Migration. I construct this database with archived passenger lists from the ships that transported emigrants abroad, which cover 73 percent of the total emigration flows in the period. The database is the most comprehensive of its kind, and I document its validity and representativeness in a number of empirical tests. Combining these two data sources allows me to observe every Scandinavian who lived and nearly everyone who emigrated in the Age of Mass Migration.

To study the relationship between collectivism and migration, I construct three proxies of collectivism, using information from the population censuses. First, I characterize people as collectivists if they give their children common first names. The proxy builds on a large literature in sociology and psychology that has tied the use of common first names to a preference for conformism and a sense of shared identity, which are core collectivist traits. Given its novelty to research in economics, I validate the proxy with comprehensive tests.[5] Using global survey data, I show that differences in name-giving today correlate with differences in self-reported collectivism – both across and within countries. Moreover, I show that the use of common first names over time correlates with more or less collectivist language use in 19th-century Swedish newspapers. Second, I construct two additional proxies that characterize people as collectivist if they live in extended family structures and if they marry someone of the same local origin. The latter two are acknowledged manifestations of a preference for living in interconnected networks, and thus collectivism.[6]

The empirical analysis of the paper consists of two parts. In the first part, I identify cultural selection in migration at the individual level. For this, I link emigrants and stayers to their childhood households in the population censuses. Analysis of the data confirms that individuals who grew up in collectivist households were less likely to

---

[5]The proxy is also adopted in Bazzi et al. (2020), who rely on the validation analysis in this paper.

[6]The preferred proxy is based on the giving of first names, which reflects a less constrained choice. The choice of marital partner is for instance impacted by the preferences of others in the marriage market.

emigrate. To address concerns that the main results capture economic rather than cultural selection in migration, I control for a host of relevant childhood characteristics. These include past migration and descent, which could impact both collectivism and emigration, socioeconomic characteristics of the parents, and factors that impact the likelihood of inheritance within the family. Cultural factors account for around 10 percent of the mean likelihood of emigration in the sample.

Not all emigrants were equally intensely selected. Some collectivists emigrated, and some individualists stayed. Examining the correlates of this variation, I show that when individuals from collectivist households emigrated, they tended to be followers rather than pioneers. They were more likely to emigrate subsequent to strong community emigration, emigrate collectively with other people from their family or town of origin, and settle with fellow Scandinavians abroad. These results support the cultural interpretation of the main selection results: Collectivist households value the ability to maintain strong ties to their existing networks. When this is feasible, the costs of migration to collectivists fall, and they become more likely to emigrate.

In the second part of the analysis, I ask if selective emigration was associated with lasting cultural change in migrant-sending districts. When emigrants select on collectivist traits, their departure causes an immediate change in the share of people with these traits in the home population. This is a purely compositional effect that I denote the *cultural shock* of emigration. If the post-migration, rather than pre-migration, share of collectivist traits persists over time, the result is lasting cultural change. The empirical analysis of the two elements in this reasoning – the shock and persistence – uses the 64 subnational Scandinavian districts as the units of observation.

First, I quantify the cultural shocks of emigration across districts and decades. Specifically, I calculate this compositional effect as the change in prevalence of collectivist traits in the home population when emigrants are included and when they are not. The size of these shocks depends not just on the magnitude of emigration but also on the intensity of selection among emigrants. In the most affected districts, selection in the Age of Mass Migration caused a 5-9 percent increase in the share of people with common first names. Within each country, the cultural shocks accounted for a considerable share of the cross-district variation in the practice of collectivist norms.

Second, I examine whether the cultural shocks were associated with persistent cultural change across districts. That is, if the compositional change in collectivist traits caused by selective emigration impacted cultural norms and beliefs in subsequent

generations. The evidence suggests that this was the case. I begin with an analysis of the evolution of norms over the Age of Mass Migration. In a dynamic panel setting that controls for district fixed effects, I document that larger cultural shocks correlated with a strengthening of collectivist norms in the following decades. I then document that the cultural change that took place in the period was sufficiently persistent to leave a long-run impact on contemporary Scandinavian culture. Cross-section evidence from Scandinavian surveys shows that collectivist beliefs today are stronger in districts that experienced larger cultural shocks in the Age of Mass Migration. Here, respondents are less likely to value individual autonomy, more likely to feel a strong connection to their local community, and more likely to trust family members over other people in society. Scandinavian districts that experienced larger outflows of particularly selected individuals in the Age of Mass Migration are thus relatively more collectivist today.

Finally, I highlight that persistence in the context of this paper was driven by inter-generational cultural transmission. I provide evidence that a share of this transmission took place within the family. Outside the family, peers and role models are likely to have played an important role as well.

Below, the contribution of this paper is put in relation to the existing literature. The next section describes the historical context and conceptual considerations. Section 3 details the main data sources and cultural proxies. Section 4 tests for cultural selection in migration using individual-level data, and Section 5 studies the implications of selection for cultural change in migrant-sending districts. Section 6 concludes.

Related Literature   The paper sheds new light on the cultural consequences of migration. To the best of my knowledge, I provide the first empirical evidence of persistent cultural implications of migration in migrant-sending places.[7] Previous research has been theoretical (Olsson and Paik, 2016; Rapoport et al., 2020) or concentrated on the contemporaneous diffusion of preferences from migrants to their original communities (Daudin et al. 2018; Tuccio et al. 2019).[8] The paper extends on this research by tracing the cultural impact of migration over the long run and by providing direct evidence on the underlying mechanism. Thereby, the paper also complements recent studies on the

---

[7]Thereby, the paper contributes more broadly to the growing literature on the effects of migration in sending places (Docquier and Rapoport, 2012; Karadja and Prawitz, 2019; Fernández-Sánchez, 2020).

[8]Olsson and Paik (2016) document a link between historical agriculture and present-day collectivism across Western regions. They theorize that agriculture fosters both collectivism and the out-migration of individualistic farmers. Rapoport et al. (2020) develop a unified framework to study cultural convergence between sending and arrival places, which they test with global contemporary survey data.

persistent cultural impact of immigration to the United States (Fernández and Fogli, 2009; Grosjean, 2014; Fouka et al., 2021). Less directly associated with migration is the study by Bazzi et al. (2020), who show that American individualism was fostered by a history of living on an expanding frontier. My findings complement theirs as I show that migrants who entered the United States at this time, to potentially settle on the frontier, were already selected in favor of individualistic traits.

The findings of the paper also improve our understanding of the determinants of migration. There is widespread documentation that economic characteristics such as skills and wealth predict emigration (Borjas, 1987; Abramitzky and Boustan, 2017). This includes Abramitzky et al. (2013), who find that Norwegian migrants in the Age of Mass Migration were negatively selected on economic prospects. However, the selection of migrants takes place on other dimensions as well. Related studies have found contemporary evidence of *cultural* selection on risk attitudes, moral values, religiosity, and gender-egalitarian attitudes (Jaeger et al., 2010; Docquier et al., 2020; Michaeli et al., 2021). An advantage of these studies is that they rely on precisely elicited measures from surveys. However, this also means that their sample sizes are modest. Together, these studies show that culture can function as a barrier to geographical mobility. In turn, this can impact the efficient allocation of labor within and across economies (Munshi and Rosenzweig, 2016).[9] I complement this research not just by focusing on another cultural dimension, but by offering more detail on the selection process with data on the universe of the population at the time. This includes disentangling cultural and economic factors in migration as well as examining patterns of selection over time and across migration situations.

Finally, the paper contributes to the literature on culture and identity in economics. This literature has provided widespread documentation that contemporary cultural differences have historical roots.[10] The majority of the evidence relies on contemporary measures of cultural differences. Therefore, the processes of cultural persistence and change remain less understood.[11] The level of empirical detail in this paper allows me to shed light on a specific channel of persistent cultural change, migration, with

---

[9]The study explains spatial wage disparities in India with lower rates of migration among individuals who benefit economically from rural insurance networks. While this documents an impact on migration of cultural structures, the underlying mechanisms are economic rather than cultural.

[10]For general reviews, see Nunn (2012) and Gershman (2017). For the roots of collectivism and individualism, see Olsson and Paik (2016), Schulz et al. (2019), Buggle (2020), and Bazzi et al. (2020).

[11]Among the notable exceptions are Voth and Voigtländer (2012), Bentzen (2019), Becker and Pascali (2019), Fouka (2020), and Giuliano and Nunn (2021).

evidence at the micro and macro level as well as over varying time horizons.

2 Background

2.1 The Age of Mass Migration

Between 1850 and 1920, more than 40 million Europeans emigrated to the New World where the majority settled in the United States (Taylor and Williamson, 1997). The Scandinavian countries experienced some of the highest emigration rates in Europe in this period. According to official statistics, displayed in Figure 1, emigration peaked in the 1880s where around 10 percent of the Norwegian and 7 percent of the Swedish populations left in a single decade. Although Danish emigration flows were weaker, more than 10 percent of the average population left altogether. Total emigration amounted to around 38 and 22 percent in Norway and Sweden, respectively – numbers only matched by Ireland and Italy (Hatton and Williamson, 1992). The rate of emigration varied considerably across historical districts, which can be seen from the map in Figure 1.[12] The main emigration ports were Gothenburg, Malmö, Oslo, Trondheim, Bergen, Copenhagen, and Hamburg.

The rise of mass emigration was spurred by the shift from wind to steam-driven ship transportation in the late 1860s. This reduced the cost of a voyage across the Atlantic to an amount that represented 18 percent of the annual earnings of a Norwegian farm laborer (Abramitzky et al., 2012).[13] Emigration was further facilitated by a near absence of regulatory migration policies in major destination countries. For example, the United States passed its first restrictive immigration law in 1917.[14]

In related empirical research, the causes of migration in this period have been linked to economic circumstances. Years of failed harvests, structural transformation, and economic inequality generated a push of particularly young people with poor economic prospects who benefitted from better job opportunities and access to cheap

---

[12]The historical districts refer to the administrative divisions immediately below the level of the state at the time: 24 Swedish *län*, 20 Norwegian *fylke*, and 20 Danish *amter*. Despite a few merges, these boundaries have remained in place in Sweden and Norway. In Denmark, the number of districts has halved. These districts form the geographical boundaries for most of the empirical analysis, except when including contemporary data. All analyses exclude the Danish districts of Aabenraa, Sønderborg, Tønder, and Haderslev, as they were under German rule until 1920.

[13]The journey length was also significantly reduced from around 30 to eight days by 1870 (Cohn, 2005).

[14]The 1917 law required a literacy test for entry into the United States. Country-specific immigration quotas were implemented in 1921 and 1924, which is after the period of interest in this paper. Likewise, Canadian immigration laws were not restrictive towards Scandinavians until late in the Age of Mass Migration. Until then, immigration laws targeted other nationality groups.

Figure 1. Overseas emigration from Scandinavia 1850-1920
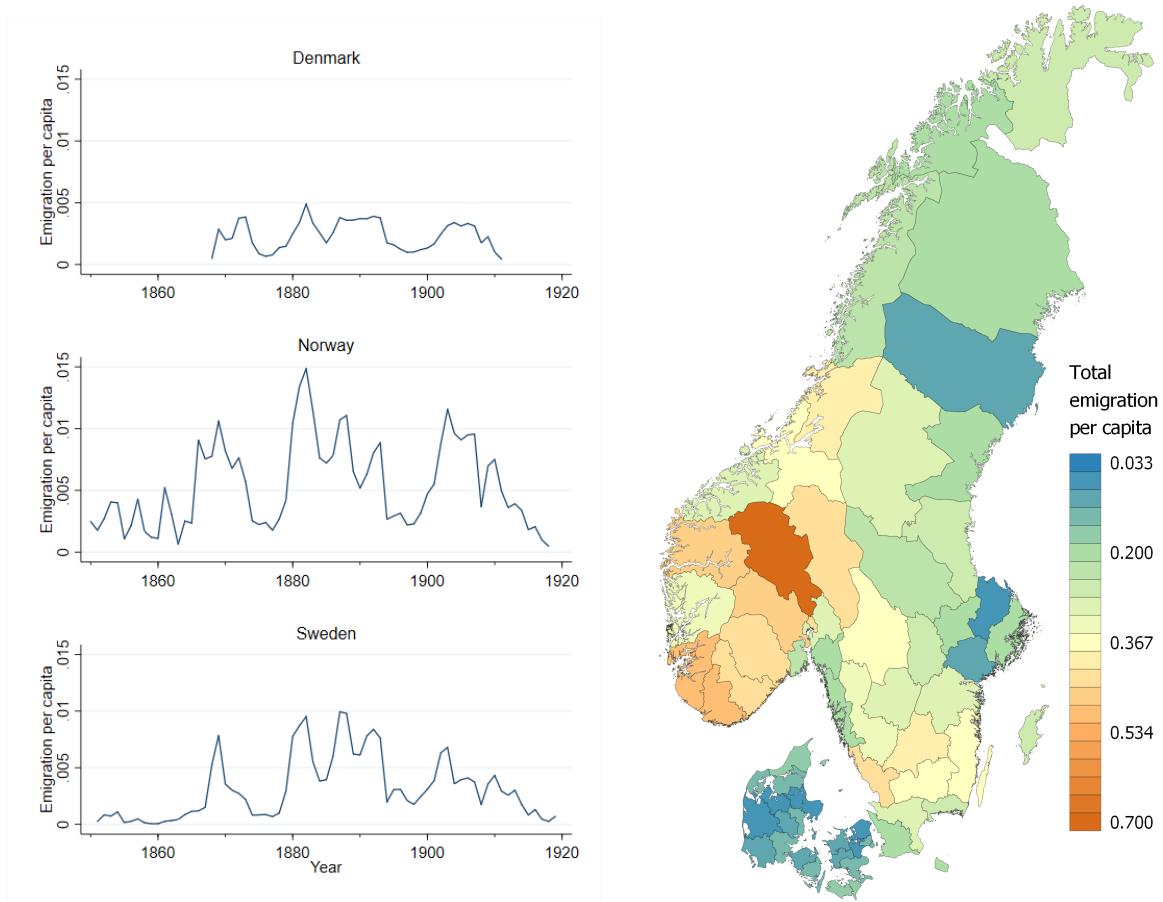
Notes: Yearly emigration per capita (left) and total emigration 1850-1920 divided by the average population (right). The average population is calculated from mid-decade population levels in the same period. Each geographical unit represents one historical district. The Swedish and Norwegian statistics are from official national accounts (NOS, 1921; SCB, 1969). The Danish statistics are aggregated from collected passenger lists available from 1868 (see Section 3 for details).

land in the New World (Hatton and Williamson, 1992; Abramitzky et al., 2012, 2013; Karadja and Prawitz, 2019).

Yet, historians agree that migrants were motivated by more than hopes of escaping poverty (Semmingsen and Haugen, 1978; Indseth, 2006; Sønnichsen, 2013). Private letters, diaries, and newspaper articles of the time reveal that ideas of personal freedom and social equality offered by American society were of great value to the migrants. In the United States, people were free to pursue their own goals. At the same time, many letters talk of the emotional difficulties related to having left their families and communities behind. This suggests that cultural factors were part of the migration decision in addition to economic incentives.

2.2   Conceptual Considerations: Selection and Cultural Change

The act of migration involves leaving familiar social networks behind. Depending on the value placed on these networks, the migration experience is potentially associated with great psychological distress (Eisenbruch, 1991; Bhugra and Becker, 2005). In social psychology, a prominent theory, the *voluntary settlement hypothesis*, proposes that people with individualistic cultural traits are more likely to migrate and settle somewhere new than their collectivist counterparts (Kitayama et al., 2006).

Individualism is commonly defined as the belief that humans constitute autonomous units in loose social networks. Here, individual identity is derived from inner attributes, abilities, and personality traits. In contrast, collectivism emphasizes the individual as embedded in larger and interconnected networks, where identity is derived from social relations. This difference in self-concepts translates into different sets of cultural values, which include the emphasis on distinction over conformity and the pursuit of individual versus collective goals (Heine and Ruby, 2010). Based on these characteristics, it seems likely that culture plays a role in the migration decision. People with individualistic traits suffer a lower cost of abandoning existing social networks – both in terms of loss of identity and support system. They may also experience personal satisfaction from doing something that requires and shows independent effort.

In its original form, the voluntary settlement hypothesis predicts that people with individualistic cultural traits are more likely to migrate, and vice versa. If migrants select, in part, on cultural traits, then migration may impact the overall evolution of cultures. Emigration must be associated with an immediate change in the prevalence of individualistic and collectivist cultural traits in the migrant-sending population. This is a simple compositional effect that I refer to as the *cultural shock of emigration*. For a shock to have persistent cultural effects, channels of intergenerational cultural transmission need to be sufficiently efficient. Theoretical and empirical research has established that individual preferences and values are passed along from parents to children and influenced by other members of society (Cavalli-Sforza and Feldman, 1981; Bisin and Verdier, 2000, 2001; Dohmen et al., 2012; Chowdhury et al., 2018).[15]

---

[15]I discuss and provide evidence of the mechanisms of cultural transmission in Section 5.3.

3 Data Sources and Cultural Proxies

3.1 Population Censuses: 1845-1901

To examine selection and cultural change in the Age of Mass Migration I rely on two main sources of individual-level data. One is a collection of complete-count historical census data on the Scandinavian populations. The data cover the years 1845, 1880, and 1901 in Denmark; 1880, 1890, 1900, and 1910 in Sweden; and 1865, 1900, and 1910 in Norway.[16] For each year and country, the censuses provide information on everyone in the population and their respective households. This includes details on birth year and location, residence, full names, occupations, and family interrelationships.

3.2 Passenger Lists: 1868-1920

My second main data source is a novel database on the majority of Scandinavians who emigrated in the Age of Mass Migration. I construct this database using archived passenger lists from the ships that transported emigrants to their destinations.

Ships provided the only means of transportation for overseas migration at the time. Ticket sales were conducted by local agencies on behalf of the shipping companies. Unauthorized ticket sales were prohibited in 1868 in Denmark and 1869 in Norway and Sweden. To maintain their authorization, agencies were required to report the personal details of all passengers to the authorities. This was part of an effort to secure passenger rights and monitor the growing migration flows (Bender, 2007).

The passenger lists have since been digitized and stored in various archives. I collect and combine these records into a database of 1.5 million first-time emigrants.[17] For each emigrant, I know their date of emigration, full name, gender, birth year, and last place of residence. For most emigrants, I also know who traveled together based on the reference number from the tickets that they bought to embark the ships.

The database is not complete. Compared to official accounts of overseas emigration,

---

[16]The collection represents every digital and complete-count population census that is available from the period. The Swedish and Norwegian censuses are available through the North Atlantic Population Project (Minnesota Population Center, 2017). The Danish censuses are obtained through the Danish National Archives. The Swedish and Norwegian censuses are of particularly high quality with numerical codes on occupations, locations, and interrelationships between household members. I assign these codes to the Danish census data based on enumerator descriptions from the time of data collection.

[17]The sources are: The Copenhagen Police Emigration Protocols 1869-1908 (DK), "Emihamn" Emigration Records 1869-1950 (SE), and regional emigration lists 1869-1930 (NO, obtained through the National Archives). Information on residence is stored in strings in the source data, which I transform into codes identical to those used in the population census records. To focus on first-time emigrants, I exclude individuals who state a foreign location as their last place of residence.

it covers 73 percent of the total emigration flows from 1868. Around 300,000 emigrated before 1868 and are not included. Comparing emigration numbers from official accounts and the passenger lists in Appendix Table D.1 reveals a strong correlation over time and space.[18] The database is thus not just the most comprehensive of its kind. It is also representative of actual emigration flows.

To further test the representativeness of the passenger list database, I compare it to the corresponding population census records. Specifically, I compare the change from one population census to the next in the number of people who share the same first name, birth year, and residence district to the number of emigrants with the same characteristics who left between these census years. The results in Appendix Table D.2 show that emigration significantly and strongly predicts actual change along these dimensions in the population. This finding speaks to the high quality of the migrant database, especially in light of unobserved effects of mortality and internal migration that hamper a clean empirical test.

3.3 Additional Data

I use a number of additional data sources. First, to study emigration behavior in the United States, I use the digitized 1900 and 1910 US census from Ruggles et al. (2021). Second, to study the long-run cultural consequences of the Age of Mass Migration in Scandinavia, I use contemporary survey data from the World Values Survey and the European Value Study. These data report the cultural attitudes of more than 13,000 Scandinavians today. Third, I use aggregate population and emigration statistics from historical national account reports.[19] Finally, I construct geographical area and distance variables at the subnational district level to be used in relevant analyses.

3.4 Proxies of Collectivism

With the available historical data, I can trace individuals from their households in Scandinavia onto the ships that carried them across the Atlantic. Studying the relationship between collectivism and migration requires cultural indicators that can be traced back in time. This is not feasible with indicators obtained from surveys or experiments, which are typically used in economic and cultural research. Instead, I construct three

---

[18]The test is restricted to Sweden and Norway, where official statistics on overseas emigration across districts are available.

[19]Norwegian statistics are from NOS (1921) and the *Kommunehefte* series at ssb.no, Swedish statistics from SCB (1969), and Danish population statistics from the Census Abstract Series available at dst.dk.

proxies of collectivism with information on *first names*, *family structures*, and *marriages* from the population censuses.

Common First Names   The main proxy of the paper characterizes people as collectivists if they give their children common first names. The proxy builds on a large literature on names in sociology and psychology (Mateos, 2013). Here, the use of common first names has been closely tied to a preference for conformism and a sense of shared identity, which are core collectivist traits.[20] Choosing a first name that is common within a community signals that this is where you belong. Vice versa, choosing an uncommon name signals independence from the surrounding communities. This balance between fitting in or standing out is regarded as one of the most important factors that new parents consider when deciding on a first name (Lieberson, 2000).

The relationship between collectivism and the use of common first names has found small-scale empirical support in a number of studies.[21] I extend on this literature in Appendix A, where I conduct a broad validation analysis across time and space. Figure 2 provides scatter plots that summarize the findings of the first part of this analysis. It shows that global differences in name-giving today correlate with differences in self-reported collectivism – both across (panel A) and within (panel B) countries. In the second part of the validation analysis, I show that the use of common first names over time correlates with more or less collectivist language use in 19th-century Swedish newspapers.

I compute the common first name proxy for each individual in the historical population censuses. First, I calculate the relative frequency of his or her name among people who were born in the previous ten birth cohorts in the same subnational district. These are the names most likely to be considered by parents when choosing a name for their children. Then, I construct the proxy as a binary variable that equals one if the first name was among the *10 most frequent* in this pool of names, and zero otherwise.[22,23] I

---

[20]Previous research has focused on trends in the use of common first names to study cultural differences across the US and Japan (Varnum and Kitayama, 2010) and to document rising individualism over time in Western countries (Lieberson and Lynn, 2003; Twenge et al., 2010).

[21]Studies on families in the United States show that parents motivate a choice of more common names with a desire for the children to fit in rather than stand out (Zweigenhaft, 1981; Emery, 2013). Surveys of children reveal that bearers of common names are more likely to value conformist behavior (Schonberg and Murphy, 1974). Similarly, Dutch families that choose more common first names for their children spend more time with other people of their local communities (Bloothooft and Groot, 2008).

[22]The focus on the ten most frequent names follows the related psychological research (Varnum and Kitayama 2010). In the empirical analysis, I test the sensitivity of this definition.

[23]Before computing the proxy, I add the names of emigrants who emigrated before the census year. Otherwise, emigrant names would appear more infrequent. All results are robust to not doing this.

Figure 2. Validation using contemporary survey data

(a) Across countries

(b) Within countries
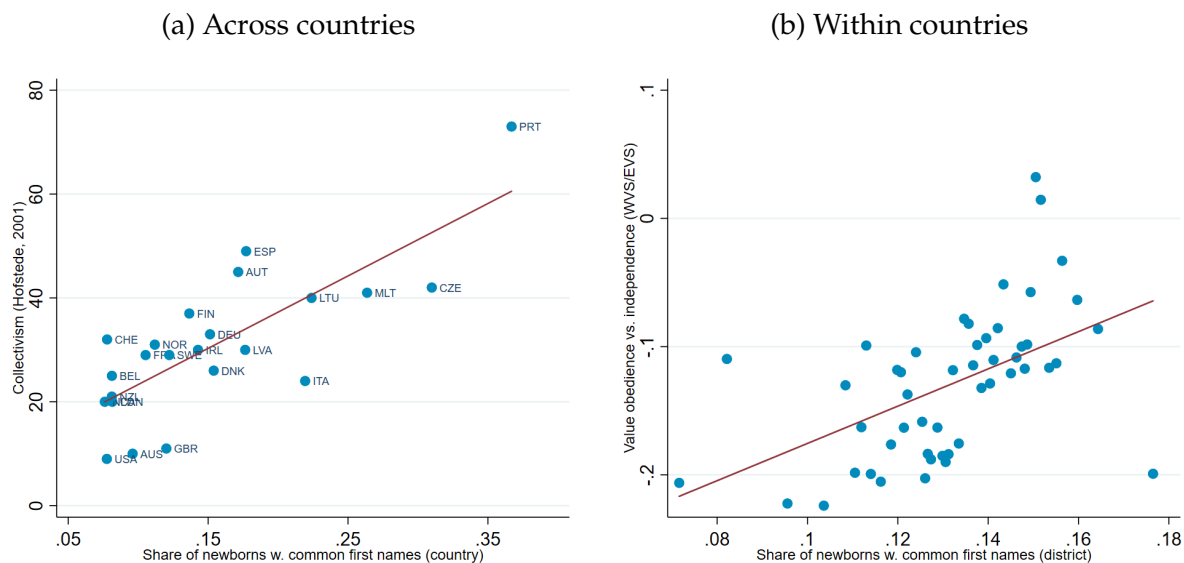
Notes: Validation of the common first name proxy using contemporary names and survey data from 23 (panel A) and 13 (panel B) Western countries. Panel A is a scatter plot of the Hofstede (2001) measure of collectivism (scale 0-100) on the y-axis and the share of newborns with common first names in 2015 on the x-axis. Panel B is a scatterplot of 50,115 respondents, binned into 50 equally sized bins, in the World and European Values Surveys with their value of *obedience* over *independence* (scale -1 to 1) on the y-axis and the share of newborns in their subnational district of residence with common first names in 2015 on the x-axis. This plot controls for country and survey wave fixed effects along with basic respondent characteristics (age, age squared, gender).

do this separately for male and female names.

Table 1 provides an overview of the ten most frequent first names among children in the earliest census from each country. Names from the most recent censuses are shown in Appendix Table D.3. It is clear that first names were highly concentrated at this point in time. Many children were given the same few names. Over time, the concentration of first names weakened. Note that the tables are for illustrating purposes. As described above, the proxy used in the rest of the paper uses naming patterns at the subnational district level to identify common first names.

A concern is that first names in the digitized censuses contain spelling mistakes. Therefore, I construct versions of the proxy based on the phonetic spelling of names or that exclude names likely to be too ridden with error. Another concern is that the proxy captures aspects that are not associated with collectivism. This includes gender, birth order, and socioeconomic status.[24] It also includes the tradition in some families

---

[24]Note that these variables may as well be correlated with collectivism. Firstborn children and boys generally receive more common first names. This has been explained by their traditional role as providers in the family, for which reason their ability to fit into society was deemed more necessary (Lieberson, 2000). Likewise, higher socioeconomic classes may have weaker collectivist traits because they are economically less vulnerable and less reliant on the communities for potential support.

Table 1. Most frequent first names among children in select censuses

Census:	Denmark 1845		Norway 1865		Sweden 1880	
Rank	Boys	Girls	Boys	Girls	Boys	Girls
1	Jens (10%)	Ane (14%)	Ole (5%)	Anne (4%)	Johan (10%)	Anna (11%)
2	Hans (10%)	Maren (8%)	Johan (4%)	Anna (3%)	Carl (9%)	Emma (4%)
3	Niels (9%)	Karen (8%)	Hans (4%)	Karen (2%)	Karl (8%)	Ida (3%)
4	Peder (6%)	Anne (4%)	Peder (2%)	Ingeborg (2%)	Gustaf (5%)	Maria (3%)
5	Anders (4%)	Marie (4%)	Carl (2%)	Maren (1%)	Axel (4%)	Johanna (3%)
6	Christian (4%)	Johanne (3%)	Lars (2%)	Marie (1%)	Anders (3%)	Hilda (3%)
7	Christen (4%)	Mette (3%)	Nils (1%)	Elen (1%)	Erik (2%)	Alma (3%)
8	Lars (3%)	Kirsten (3%)	Anders (1%)	Hanna (1%)	Per (2%)	Hulda (3%)
9	Rasmus (3%)	Else (1%)	Andreas (1%)	Johanne (1%)	Nils (2%)	Augusta (2%)
10	Søren (3%)	Caroline (1%)	Karl (1%)	Ane (1%)	Frans (2%)	Hilma (2%)

Notes: The most frequent first names among girls and boys aged 0-10 in the earliest census from each country. The shares presented in the parentheses are calculated relative to all girls and boys separately.

to pass along the names of older family members. Moreover, although Scandinavia was considered extraordinarily homogenous in terms of ancestry and shared history, aspects of diversity may still impact the choice of first names. Fortunately, these potential confounders are observable in the censuses and can be included as control variables in the empirical analysis. See Section 4.3 for details.

Additional Proxies   A *second* proxy characterizes a family as collectivist if it has an extended, rather than nuclear, structure. The extended family refers to the co-residence of family members beyond parents and their children, who constitute the nuclear family. I define a family as extended if it includes a parent to another adult family member. A *third* proxy characterizes a married couple as collectivist if both parts are born in the same subnational district. At this time, the subnational district likely represented the outer boundaries of an individual's social group. I refer to this as intra-marriage.

The extended family structure and intramarriage are acknowledged manifestations of a collectivist culture (Macfarlane, 1978; Vandello and Cohen, 1999). They describe behavior that is in accordance with a preference for living in interconnected networks with strong social ties. A drawback is that they are the result of constrained choices. The choice of marital partner is, for instance, impacted by the preferences of others in the marriage market. Therefore, my main proxy is based on the giving of first names, which better reflects a pure choice.

4 Cultural Selection in Migration

4.1 Linked Samples

To detect selection in migration, I link a subsample of emigrants and stayers to their childhood households. The aim is to compare childhood circumstances that make migration more or less likely. Focusing on circumstances in childhood rather than adulthood is particularly well-suited for the objective of this paper. First, childhood circumstances are pre-determined from the perspective of the individual who decides to emigrate or stay later in life. Second, they are likely predictive of this decision because they shape the economic and cultural prospects of the individual.

My linking method follows related studies (Abramitzky et al., 2013; Eriksson, 2015; Fouka, 2020). I start with a list of males who towards the end of the Age of Mass Migration had either emigrated or stayed in Scandinavia. Emigrants are identified from the passenger lists and stayers from the most recent census in each country (1901 in Denmark, 1910 in Norway and Sweden). From this list, I extract a subsample of males who were below the age of 15 in the earlier censuses. I then link these individuals to their younger selves based on their name, birth year, and district of origin. Females are left out due to the custom of surname change at marriage, which makes them difficult to link across data sources. My baseline linking method results in a sample of 700,000 children that represent 30 percent of the linkable child population. Further details can be found in Appendix B.1.

Given that I can only link a subsample of emigrants and stayers to their childhood households, a concern is that the sample is not representative. This is confirmed in Table 2 that reports the summary statistics for the linked sample and balance tests compared to the total child population. Children with common first names are, for instance, underrepresented in the sample. Given the sample sizes, there are significant differences in other variables too, but they are economically small. I deal with this in two ways. First, by linking emigrants and stayers *jointly*, I minimize the relative unrepresentativeness between these groups. Second, I conduct a group-level analysis in Section 4.3 that includes the entire sample of emigrants and stayers.

4.2 Individual-Level Results

I estimate the impact of collectivism on migration with the following model:

Table 2. Summary statistics and balance tests

	Linked sample		Population		Difference	
	Mean	S.D.	Mean	S.D.	Value	S.E.
Emigration	0.110	0.313				
Out-migration (only stayers)	0.120	0.325				
Common first name	0.415	0.493	0.549	0.498	-0.134	(0.000670)
Common last name	0.455	0.498	0.589	0.492	-0.134	(0.000666)
Parents intra-marriage	0.776	0.417	0.781	0.414	-0.005	(0.000591)
Extended family	0.061	0.239	0.064	0.244	-0.003	(0.000328)
Age	6.224	3.908	6.480	3.967	-0.256	(0.00533)
No. of brothers	1.691	1.375	1.577	1.333	0.114	(0.00181)
Firstborn child	0.409	0.492	0.449	0.497	-0.040	(0.000669)
Birth order	2.061	1.178	1.942	1.114	0.119	(0.00152)
Foreign parent	0.009	0.096	0.013	0.112	-0.003	(0.000147)
Moved since childhood	0.951	0.216	0.947	0.224	0.004	(0.000300)
Family size	6.344	2.060	6.172	2.015	0.172	(0.00273)
No. of first names	1.667	0.591	1.651	0.613	0.015	(0.000820)
Name shared by family member	0.184	0.388	0.253	0.435	-0.069	(0.000573)
Related to household head	0.988	0.108	0.983	0.131	0.006	(0.000170)
Urban residence	0.185	0.389	0.165	0.371	0.020	(0.000505)
Servants in household	0.191	0.393	0.199	0.399	-0.008	(0.000536)
Parent on poor relief	0.014	0.119	0.017	0.128	-0.002	(0.000170)

Notes: The table includes males below the age of 15 in the censuses DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, and SE-1900. Linked N=691,969; Population N=2,701,758.

$$M_{idyc} = \alpha_p p_i + \phi_d + \phi_y + \phi_c + \alpha_X X_i + \varepsilon_{idyc}, \tag{1}$$

where the unit of observation is an individual male i born in district d and decade y, who is observed as a child (below the age of 15) in any of the early censuses c. M_{idyc} is an indicator that equals one if the individual emigrated and zero if he stayed. p_i is one of three proxies of collectivist norms practiced in the childhood household. ϕ_d, ϕ_y, and ϕ_c denote birth district, birth decade, and census fixed effects. X_i is a set of childhood control variables. ε_{idyc} is the error term.

A negative α_p means that emigrants came from less collectivist households than the rest of the population. To the extent that children inherit the norms of their parents, this implies that the emigrants themselves were less collectivist. This is supported by the literature on transmission between parents and children discussed in Section 2.2 on which I present supportive evidence in Section 5.3. Of course, the validity of this interpretation also depends on controlling for potential confounders.

Table 3 reports the results from estimating Model 1 on my baseline sample. Standard errors are clustered at the birth district × birth decade × census level. The proxy of collectivism differs across the panels of the table. In Panel A it is the indicator of

having a common first name, in Panel B the indicator of living in an extended family structure, and in Panel C the indicator of intra-group marriage between the individual's parents. The estimated coefficient on these proxies is negative in all specifications. This implies that males who grew up in collectivist households were on average less likely to emigrate later in life.

The evidence on cultural selection in migration does not appear to be confounded by other relevant childhood circumstances. This includes family demographics such as sibling structure and measures of past migration, which I control for in Column (2).[25] It also includes other characteristics of the individual's name: In column (3) I control for the number of recorded first names along with indicators of having a common last name and a first name that is phonetically identical to that of an older family member. These indicators capture local descent, which could impact both collectivist norms and emigration, and the strength of family traditions.[26]

Finally, column (4) controls for socioeconomic characteristics of the family measured by urban residence, having servants employed, and of any parent receiving poor relief. Here, the coefficients in panels B and C drop but remain negative and significant. To further disentangle cultural and economic factors in migration, I include father occupation fixed effects in a robustness test in Appendix Table D.8.

The evidence in this section supports the hypothesis of collectivist selection in migration. Moreover, the coefficients are sizable. Depending on the cultural proxy, growing up in a collectivist household is associated with a 0.8-1.1 percentage point lower propensity to emigrate in the full specification. These numbers account for a large share of the mean likelihood of emigration in the sample, which is 11 percent. The coefficients are also comparable to those on other relevant variables such as being the firstborn child or having servants employed, which are associated with a 0.4 and 1.6 percentage point lower propensity to emigrate. These variables proxy the likelihood of inheritance and family wealth, which have been the focus of previous research (Abramitzky et al., 2013).[27]

---

[25]These include the age of child and parents, log family size, log number of brothers, log birth order, and indicators for being the firstborn, the parents being foreign, or having moved since one's birth.

[26]The common last name indicator is calculated similarly to the common first name proxy.

[27]Coefficients on control variables are not reported in the paper, but they are available upon request.

Table 3. Collectivist norms and emigration

Dependent var.	Emigration			
	(1)	(2)	(3)	(4)
Panel A				
Common first name	-0.0147***	-0.0137***	-0.0120***	-0.0114***
	(0.002)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)
Observations	691,969	691,969	691,969	691,969
R squared	0.14	0.15	0.15	0.15
Panel B				
Extended family	-0.0117***	-0.0105***	-0.0103***	-0.0079***
	(0.002)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)
Observations	691,969	691,969	691,969	691,969
R-squared	0.14	0.15	0.15	0.15
Panel C				
Parents intra-marriage	-0.0129***	-0.0120***	-0.0122***	-0.0098***
	(0.002)	(0.002)	(0.002)	(0.001)
Mean (sd) of DV	0.108 (0.310)	0.108 (0.310)	0.108 (0.310)	0.108 (0.310)
Observations	618,934	618,934	618,934	618,934
R-squared	0.14	0.14	0.14	0.14
Demography and past migration	N	Y	Y	Y
Additional name characteristics	N	N	Y	Y
Socioeconomic status	N	N	N	Y

Notes: OLS results. The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses (DK 1845-80, NO 1865-1900, SE 1880-1900), who are linked to either the passenger lists or the most recent census for each country (DK-1901, NO-1910, SE-1910). The dependent variable equals one if linked to the passenger lists (emigrant) and zero if linked to the most recent census (stayer). Explanatory variables proxy collectivist norms in the childhood household. All regressions include census, birth decade, and birth district fixed effects. *Demography and past migration* controls include age of child, log family size, log no. of brothers, a firstborn indicator, log birth order, and indicators of foreign-born parents and of living outside one's district of birth. *Additional name char.* include, along with number of first names, indicators of having a common last name and a first name that is phonetically identical to that of an older family member. *Socioeconomic status* of the family include indicators of urban residence, having servants employed, being related to the household head, and of any parent receiving poor relief. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.3 Robustness

Group-Level Results As discussed in Section 4.1, the individual-level results may be biased by the fact that I cannot link every individual to his childhood household. Although I use different strategies to minimize this bias, concerns remain.[28] In Appendix C, I therefore carry out a group-level analysis with *all* emigrants and stayers. This includes females, who were left out of the individual-level analysis.

---

[28]For example, Appendix Table D.6 shows that the main selection results are not due to a particular aspect of my linking method.

The unit of observation is now a group of individuals who share the same first name, gender, birth year, and district of residence. For each group, I calculate their propensity to emigrate by dividing the number of migrants in the passenger lists with the number of people in the census who share the same characteristics. Cultural proxies and control variables equal averages across the individuals in the group. The analysis reveals that individuals in groups that were more likely to have grown up in collectivist households were less likely to emigrate. Results from the most representative sample imply that individuals with common first names were 1.2 percentage points less likely to emigrate. This is very much in accordance with the main results in Table 3. In the Appendix, I also assess *how* common a first name has to be in order to be associated with lower emigration rates. This evidence shows that cultural selection in migration was a broad and significant phenomenon in the Scandinavian populations.

Measurement of Common First Names Results in Table 3, Panel A, are robust to using alternative measures of commonness as seen in Appendix Table D.10. This is true for different kinds of spelling, including the phonetic equivalent and abbreviations of the name. Rare names, which are shared by less than 10 individuals in a census, may be dropped. So may names that per laws in each country are not among the officially approved names today. Instead of using the first name of the individual child, I can use that of the oldest brother in the household. The associated significant result supports the interpretation that the cultural determinants of emigration were a family rather than individual child name effect.

Additional Tests Appendix Tables D.7 and D.8 show that the results are robust to using probit estimation and to including additional fixed effects for the birth municipality, first name, last name, father's occupation, and exact birth order. A final set of robustness results appear in Appendix Table D.9. Results hold when excluding children that emigrated before the age of 15, which rules out that parents gave their children less common names while planning their own emigration. I also analyze the decision to emigrate and out-migrate separately. Out-migrants are defined as children who in the most recent census lived outside their district of birth. Results show that there was negative, but weaker, selection on collectivist norms in out-migration, which is expected given that it likely entailed smaller social costs than emigration.

4.4 Selection among Pioneers and Followers

The paper has so far documented selection in migration in a cross-section. But not all emigrants were equally strongly selected. In this section, I study the determinants and implications of this variation. Consistent with what one would expect, emigrants from collectivist households tended to be *followers* rather than *pioneers*: They were more likely to emigrate subsequent to strong community emigration, emigrate collectively, and settle with familiar networks abroad.

These results not only shed light on the dynamic process of emigration. They also support the cultural interpretation of my selection results. Collectivist households value the ability to maintain strong ties to their existing networks. When this is feasible in migration, social costs fall, and they become more likely to emigrate.

Community migration and selection intensity   To provide the first piece of evidence on pioneer and follower emigration, I examine whether the intensity of selection in migration varies over time as community emigration accumulates. The expectation is that higher past community emigration, which entails larger emigration networks abroad, reduces the social costs of emigration and weakens selection.

To study if this is the case, I create a panel over the decades 1860 to 1910 from the linked sample of emigrants and stayers. I code a new dependent variable, *emigration status*, that varies over time. For emigrants, it equals one in and after the decade of emigration, and zero otherwise. For stayers it equals zero in all decades. I then estimate a version of Model 1, where I regress emigration status on the *interaction* between time-invariant individual-level characteristics and the time-varying sum of past decennial emigration rates in the district of residence. I include a full set of control variables, add decade fixed effects, and control for the log population and age of the individual.[29]

Results in Table 4 show that cultural selection in migration was weaker in decades and districts with higher accumulated community emigration. This is implied by the negative coefficients on the cultural proxies and the positive coefficients on the interaction term. When community emigration accumulates, the sum of these coefficients approaches zero. This is in accordance with the expectation.

---

[29]Specifically, the model is given by: $M_{idyct} = \alpha_p p_i \times m_{dt} + \phi_d + \phi_y + \phi_c + \phi_t + \alpha_X X_{it} + \varepsilon_{idyct}$, where M_{idyct} is the emigration status in decade t of individual male i (born in district d and decade y and observed in census c). p_i proxies collectivist norms in the childhood household. This is interacted with m_{td}, which is the sum of past decennial emigration rates in the district of residence. ϕ_d, ϕ_y, ϕ_c, and ϕ_t are birth district, birth decade, census, and decade fixed effects. X_{it} denote control variables. Time-invariant control variables are interacted with m_{td} as well.

Table 4. Community emigration and selection intensity 1860-1910

Dependent var.	Emigration status					
	(1)	(2)	(3)	(4)	(5)	(6)
Collectivist norm	-0.0129***	-0.0114***	-0.0102***	-0.0075***	-0.0177***	-0.0130***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Past emigration	0.2550***	0.2516***	0.2622***	0.2505***	0.2133***	0.2190**
	(0.062)	(0.086)	(0.061)	(0.087)	(0.060)	(0.088)
Collectivist norm × past emigration	0.0348***	0.0349***	0.0286**	0.0235*	0.0638***	0.0480***
	(0.008)	(0.008)	(0.013)	(0.013)	(0.011)	(0.009)
Control variables	N	Y	N	Y	N	Y
Collectivist norm	Common first name		Extended family		Parents intra-marriage	
Mean of DV	0.0693	0.0693	0.0693	0.0693	0.0688	0.0688
SD of DV	0.254	0.254	0.254	0.254	0.253	0.253
Observations	1,905,151	1,905,151	1,905,151	1,905,151	1,770,402	1,770,402
R-squared	0.15	0.15	0.15	0.15	0.15	0.15

Notes: OLS results from a panel of linked male children who are observed in the decades 1860-1910. The dependent variable equals one in and after the decade of emigration and zero before emigration. For non-emigrants it equals zero in all decades. Explanatory variables proxy collectivist norms in the childhood household. *Past emigration* equals the sum of past, decennial emigration rates in the district of residence. All regressions include decade, census, birth decade, and district fixed effects. Controls include age and age squared in the last year of the decade and log district population. Individual-level controls in columns (2), (4), and (6) include demographics of the family, past migration, socioeconomic status, and additional names characteristics (detailed in Table 3). Standard errors are clustered at the census × birth decade × district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Independent or collective emigration Another way to study pioneer and follower emigration, is to study *how* people emigrated. Here, the expectation is that individuals from collectivist households tended to emigrate collectively with people from their social networks. The passenger lists provide information on three dimensions of collective emigration. First, emigrants are noted as emigrating with family or not. Second, I observe if they emigrated on the same day as other people from their hometowns. Third, ticket numbers are available for a subsample of the emigrants, which reveals if they emigrated on a joint fare that covered the passage across the Atlantic for multiple individuals.

Cross-section analysis of the entire sample of emigrants in the passenger lists shows that emigrants with common first names were more likely to emigrate collectively. This is in accordance with the expectation. Results are available in Appendix Table D.13

and they are robust to alternative calculations of the common first name proxy.[30]

Settlement in the United States More than two-thirds of the Scandinavian migrants settled in the United States. Using data from the 1900 and 1910 American population censuses, I can observe the settlement behavior of more than 60,000 of these migrants. I identify Scandinavian migrants as individuals who were born in Denmark, Norway, or Sweden, and I use their first names to proxy for growing up in collectivist households in Scandinavia.[31]

Cross-section results in Appendix Table D.15 show that Scandinavian immigrants with first names that were common in their birth countries tended to settle with fellow Scandinavians after arriving in the United States. They were more likely to reside in counties inhabited by and marry people of their own nationality. They were also more likely to choose more distinctively Scandinavian sounding first names for their children. Calculated as the popularity of a first name among Scandinavian immigrants relative to its popularity among non-Scandinavians, this outcome variable proxies ethnic identity (Fryer and Levitt, 2004; Fouka, 2020). These results confirm expected migrant behavior at the destination. Finding themselves in a new economic, institutional, and social setting, migrants from collectivist households joined familiar networks with fellow Scandinavians instead of building new networks with people of other nationalities.[32]

5 Cultural Change in Migrant-Sending Districts

The previous section showed that Scandinavians with collectivist traits were less likely to emigrate in the Age of Mass Migration. In this section, I study whether selective emigration was associated with lasting cultural change in migrant-sending districts.

When emigrants select on collectivist traits, their departure causes a compositional change in the share of people with these traits in the home population. I denote this

---

[30]Common first names are identified for each emigrant in the earliest available census relative to their birth year, gender, and last district of residence. Results are robust to using the census closest to adult age (40 years), which allows for emigrants to have moved across districts since childhood. They are also robust to defining common first names from their phonetic spelling and within a ±1 year band around the reported year of birth, to allow for discrepancies (see Appendix Table D.14).

[31]Immigrants are identified in the 1900 5% and 1910 1% IPUMS samples for the United States (Ruggles et al., 2021). I can only observe the immigrants' country, and not the district, of birth. Using the Scandinavian censuses, the common first name proxy is calculated as the share of people in the birth country with the same first name, gender, and birth year, whose name was common in their district of birth.

[32]The results are robust to using the phonetic spelling of names, as seen in Appendix Table D.16. This is relevant for immigrants who Americanized the spelling of their names. They are also robust to defining common names within a ±1 year band around the reported year of birth.

change as the *cultural shock* of emigration. If the shock persists over time, the result is lasting cultural change. In the following, I assess the magnitude of these cultural shocks and their persistence over the short, medium, and long run. I also discuss channels of persistence with a focus on intergenerational cultural transmission.

5.1 Cultural Shocks of Emigration across Districts

I define a cultural shock as the change in prevalence of collectivist traits in the population that is directly caused by the emigration of selected individuals. Formally this is given by:

$$\Delta p = \frac{P^s}{N^s} - \frac{P^s + P^e}{N^s + N^e} \, ,$$
(2)

where N denotes the number of individuals and P the number of individuals with collectivist traits who stay (^{s}) or emigrate (^{e}).

I calculate the shocks for each district and decade in the Age of Mass Migration. I proxy P with the number of individuals with common first names. Accordingly, the shock measures the increase in the share of common first names in the home population due to emigration. Since names reflect parents' choices, a shock can be interpreted as a change in the share of people with a collectivist cultural upbringing.

The calculations rely on the entire universe of census and passenger lists data. For each decade and district, I measure $\frac{P^s}{N^s}$ among people who were present – the stayers – in the final year of the decade. I measure $\frac{P^s + P^e}{N^s + N^e}$ by including emigrants from the passenger lists who left in the same decade. The measures contain some noise. For example, I use interpolation to approximate the distribution of cohorts and first names in the staying population in the years not covered by the population censuses. Further details can be found in Appendix B.2, where I also describe alternative calculations intended for robustness checks.

Figure 3, Panel A, illustrates the sum of cultural shocks over the decades of the Age of Mass Migration. To understand what this measures, assume that each shock was perfectly persistent. The sum of shocks would then equal the total change over the period in the share of people with common first names due to selective emigration. In Panel B the sum of shocks is divided by the mean share of people with common first names over the period. The most affected regions were Finnmark in Norway and Värmland and Kalmar in Sweden. Here, the share of common first names increased

by more than 3 percentage points due to selective emigration, corresponding to 5-9 percent of the mean level.

Figure 3. Sum of cultural shocks 1860-1920 across districts

(a) Absolute compositional change   (b) Relative compositional change
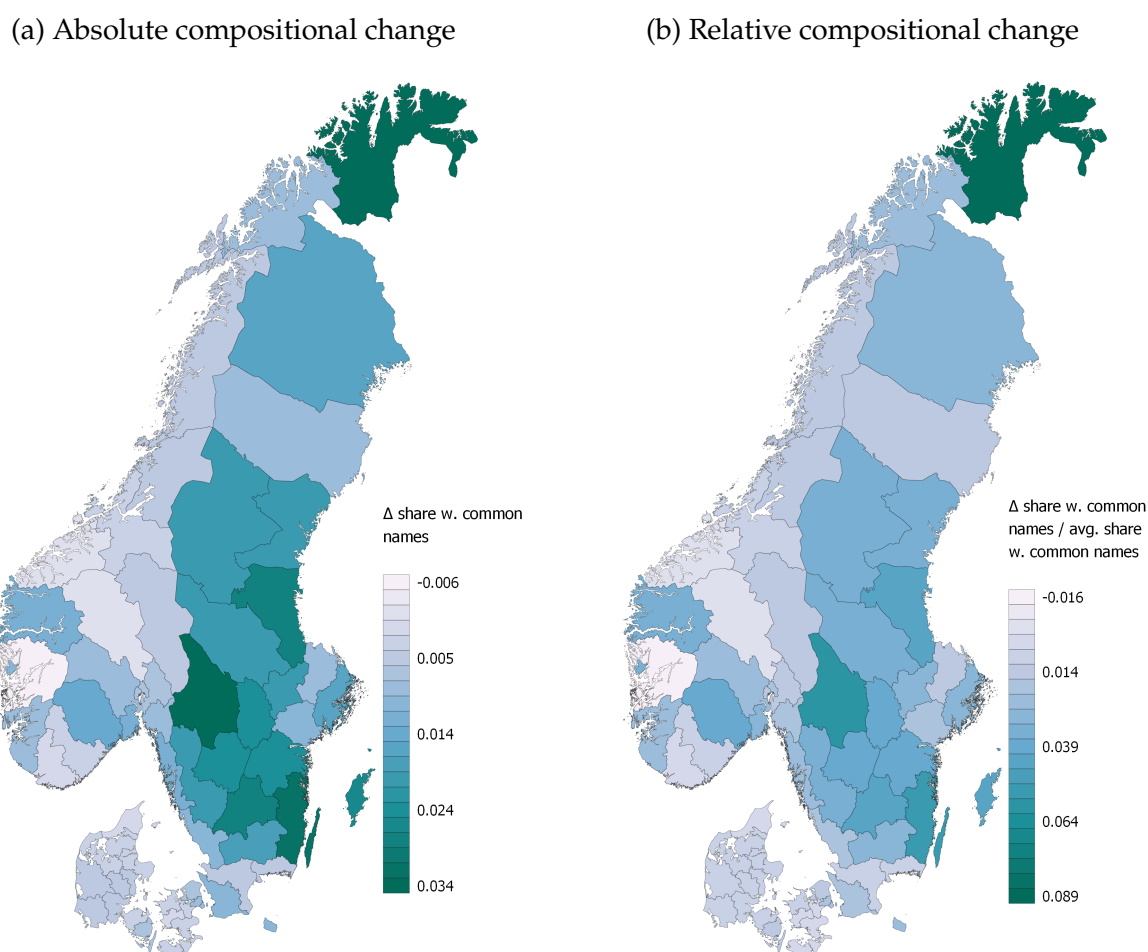
Notes: Panel A illustrates the sum of cultural shocks over the decades of the Age of Mass Migration. The shocks are calculated according to Eq. 2 as the increase in the share of the district population with common first names, which is caused by the emigration of selected individuals. In Panel B this is divided by the average population share with common first names over the period. Darker colors imply an increase in the concentration of people with a collectivist cultural upbringing, due to selective emigration. The underlying data can be found in Appendix Table D.17.

Overall, Denmark experienced the weakest impact, which makes sense given the lower rates of emigration.[33] Note, however, that the size of the shocks depends not just on the number of emigrants but also on the intensity of their selection. Since selection varied across districts, there is no strong correlation between the total rate and cultural shock of emigration as seen in Appendix Figure D.1.

The cultural shocks may appear modest in size. However, they account for a large share of the within-country variation in the practice of collectivist norms over the period. In Denmark, Norway, and Sweden, the sum of shocks in the average district

---

[33]Four districts in Norway experienced mildly negative shocks, which according to evidence in Section 4.4 may be explained by circumstances that allowed for collective emigration.

accounted for 6, 17, and 38 percent of the cross-district variation in the use of common first names, respectively (see Appendix Table D.17).

5.2 Persistence of Cultural Shocks

The next question is if the cultural shocks of emigration persisted over time. That is, if the compositional change in cultural traits due to emigration impacted cultural norms and beliefs in subsequent generations. To answer the question empirically, I begin with an analysis of the evolution in norms within the period of the Age of Mass Migration. Then, I ask if the shocks were sufficiently profound to leave a long-run impact on contemporary Scandinavian culture.

In the Short Run To study persistence in the short run, I construct a panel that spans the decades of the Age of Mass Migration. For this, I use the 1901 census in Denmark and the 1910 censuses in Norway and Sweden. I proxy the practice of collectivist norms over time with the use of common first names across birth decades. Using within-census variation like this is feasible because first names reflect the norms at the time of name-giving.[34] I combine the data with measures of the cultural shocks derived above. Then, I run panel regressions of the form:

$$p_{dt} = \beta_s s_{dt-1} + \sum_{j=1}^{2} \beta_j p_{dt-j} + \beta_X X_{dt} + \gamma_d + \gamma_t + \varepsilon_{dt} , \qquad (3)$$

where p_{dt} is the share of people born in district d and decade t with common first names. s_{dt-1} is the lagged cultural shock of emigration. X_{dt} is a set of control variables. γ_d denotes a full set of district fixed effects, which absorbs the impact of any time-invariant district characteristics, and γ_t denotes a full set of decade fixed effects. ε_{dt} is the error term.

The model includes two lags of the dependent variable to control for its dynamics.[35] Positive β_j's indicate that collectivist norms persist over time, irrespective of the impact of selection in migration in decade $t-1$. The main coefficient of interest is β_s. A positive β_s implies that an increase in the share of people with a collectivist upbringing, due to emigration, was associated with a strengthening of collectivist norms in the following

---

[34]Like everywhere else in the paper, names measures have been adjusted to include emigrants who left before the census year. Due to mortality, the measures may, however, be imprecise for earlier decades.

[35]Using just one lag of the dependent variable, second-order autocorrelation is sufficiently strong to invalidate the GMM results in Table 5.

decade.

Table 5 reports the results from estimating the model with three distinct methods: OLS without district fixed effects in columns (1)-(3), fixed effects in columns (4)-(6), and difference GMM in columns (7)-(9). The latter is to adjust for potential Nickell bias. I focus on the decades between 1860 to 1910, for which I have both passenger list and census data.[36] All regressions control for the log number, gender ratio, and mean number of first names of the people who were born in the decade and district. In columns (3), (6), and (9), I add a second lag of the cultural shock variable to account for potential delayed effects on the practice of collectivist norms. Sums of the coefficients on the lagged variables are reported in italics.

There are two findings of interest in Table 5. First, collectivist norms were highly persistent over the period. This is evident from the positive coefficients on the lagged dependent variable. Second, and as expected, the coefficients on the lagged cultural shock variable are positive and significant. This also holds in columns (2), (5), and (8) when controlling for the log of total emigration. This control variable is likely endogenous to past collectivist norms, but its inclusion limits concerns that the β_s's capture the overall impact of emigration rather than its compositional cultural impact, as intended.

Taking the coefficients at face value implies that a 1 percentage point increase in the share of people with a collectivist upbringing, due to selective emigration, was associated with a 1.02-1.30 percentage point increase in the share of people who practiced collectivist norms in the following decade. The result is weaker when accounting for district fixed effects, which are demanding in a regression with just 5 periods. Adding a second lag of the cultural shock variable in columns (3), (6), and (9) shows that the overall impact of a cultural shock requires time to play out. The sum of coefficients is significant at the 1 percent level across estimation methods. Appendix Table D.18 shows that the results are robust to alternative calculations of the cultural shock variable.[37]

The evidence presented here implies that the cultural shocks were persistent over the short run, generating relative cultural change towards stronger collectivist norms.

---

[36]The lags go further back than 1860. In the GMM estimation, I use all lags (going back to 1800) of the dependent, explanatory, and control variables as internal instruments. For the decades up to 1860, I set the shock variable equal to zero, which is justified by the very low emigration in this period.

[37]The results are also robust to controlling for the prevalence of common last names and to including a third lag of the dependent and cultural shock variable. Results are available upon request.

Table 5. Persistence of cultural shocks in the short run, 1860-1910

Dependent var.	Share of newborns with common first names								
Estimation method	OLS without district FE			Fixed effects			Difference GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Common first names, first lag	1.182***	1.178***	1.168***	0.926***	0.909***	0.904***	1.052***	1.051***	1.029***
	(0.067)	(0.067)	(0.068)	(0.064)	(0.064)	(0.063)	(0.062)	(0.061)	(0.058)
second lag	-0.293***	-0.289***	-0.280***	-0.272***	-0.283***	-0.282***	-0.338***	-0.345***	-0.338***
	(0.063)	(0.063)	(0.064)	(0.059)	(0.058)	(0.061)	(0.052)	(0.051)	(0.053)
Cultural shock of emigration, first lag	1.396***	1.296**	1.079*	1.069*	1.020*	1.052*	1.028*	1.102**	1.000*
	(0.523)	(0.521)	(0.587)	(0.611)	(0.574)	(0.558)	(0.539)	(0.559)	(0.549)
second lag			0.226			0.414			0.572
			(0.367)			(0.327)			(0.379)
Common first names, sum of lags	0.890***	0.889***	0.888***	0.654***	0.626***	0.622***	0.713***	0.706***	0.692***
	(0.017)	(0.016)	(0.016)	(0.044)	(0.045)	(0.045)	(0.038)	(0.039)	(0.036)
Cultural shock of emigration, sum of lags			1.305***			1.466***			1.572***
			(0.445)			(0.451)			(0.540)
Baseline controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Log emigration	N	Y	Y	N	Y	Y	N	Y	Y
AR2 test *p*-value							0.790	0.798	0.765
Mean (sd) of DV	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)	0.442 (0.115)
No. of districts	64	64	64	64	64	64	64	64	64
Observations	320	320	320	320	320	320	320	320	320
R-squared	0.98	0.98	0.99	0.96	0.96	0.96			

Notes: Dynamic panel results. The dependent variable is measured across 64 subnational districts in the decades 1860 to 1910. *Cultural shock of emigration* measures the increase in the share of common first names in the district population due to selective emigration. All regressions include decade fixed effects. Baseline controls include the log number, gender ratio, and mean number of first names of people born in the decade and district. *Log emigrants* enters with the same lags as the *Cultural shock* variable. The panels contain no unit roots (Levin-Lin-Chu *t*-stat -9.420, *p*-value 0.000). Robust standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In the Medium Run   An alternative to the dynamic analysis is to consider persistence over the entire Age of Mass Migration in a cross-district setting. This may appear as an inferior analysis as I cannot involve district fixed effects. The advantage, however, is that I can study the evolution in all three proxies of collectivism. Therefore, I estimate the following model:

$$p_d^{post} = \beta_s s_d + \beta_p p_d^{pre} + \beta_X X_d + \varepsilon_d , \qquad (4)$$

where p_d^{pre} and p_d^{post} denote the intensity of a collectivist norm in district d at the start and towards the end of the Age of Mass Migration, respectively.[38] The coefficient β_p captures persistence in collectivist norms over most of the Age of Mass Migration. s_d is the sum of cultural shocks over the decades between the *pre* and *post* years. Again, a positive β_s indicates that a change in the composition of people with different cultural upbringings, due to emigration, is associated with a change in the practice of cultural norms. X_d is a set of control variables and ε_d is the error term.

Figure 4 summarizes the main results. It shows the AV-plots of coefficient β_s for each collectivist norm with controls for initial population levels and total emigration, both in logs. The results suggest that the cultural shocks of emigration persisted over the medium run in the sense that they were associated with significant changes in collectivist norms. The underlying regression results and robustness tests can be found in Appendix Table D.19 and D.20.[39]

In the Long Run   Now I turn to study if the cultural shocks associated with the Age of Mass Migration were sufficiently profound and persistent to leave a long-run imprint on contemporary cultural differences in Scandinavia. For this, I use of individual-level data from Scandinavia in the World Values Survey (WVS) and the European Values Study (EVS) for the years 1996-2018.

I construct four dependent variables to capture collectivism in the WVS and EVS. First, I make use of a set of questions that ask parents about qualities that they regard as important in children. I construct a variable, ranging from -1 to 1, that takes the men-

---

[38]The *pre* values are taken from the earliest censuses (1845 in Denmark, 1880 in Sweden, and 1865 in Norway), and the *post* values are from the latest censuses (1901 in Denmark and 1910 in Sweden and Norway). The ideal census years would have been 1850 and 1920, which marks the official beginning and end of the Age of Mass Migration, but this is not feasible with the available historical data.

[39]Except for intra-marriage, the results are robust to including more control variables such as country fixed effects and geographical variables. They are also robust to using alternative calculations of the cultural shock variable. The lack of robust results on intra-marriage may be explained by this being a weaker proxy, which is impacted by e.g. internal migration patterns.

Figure 4. Persistence of cultural shocks in the medium run, AV-plots

(a) Common first names

(b) Extended families

(c) Intra-marriage
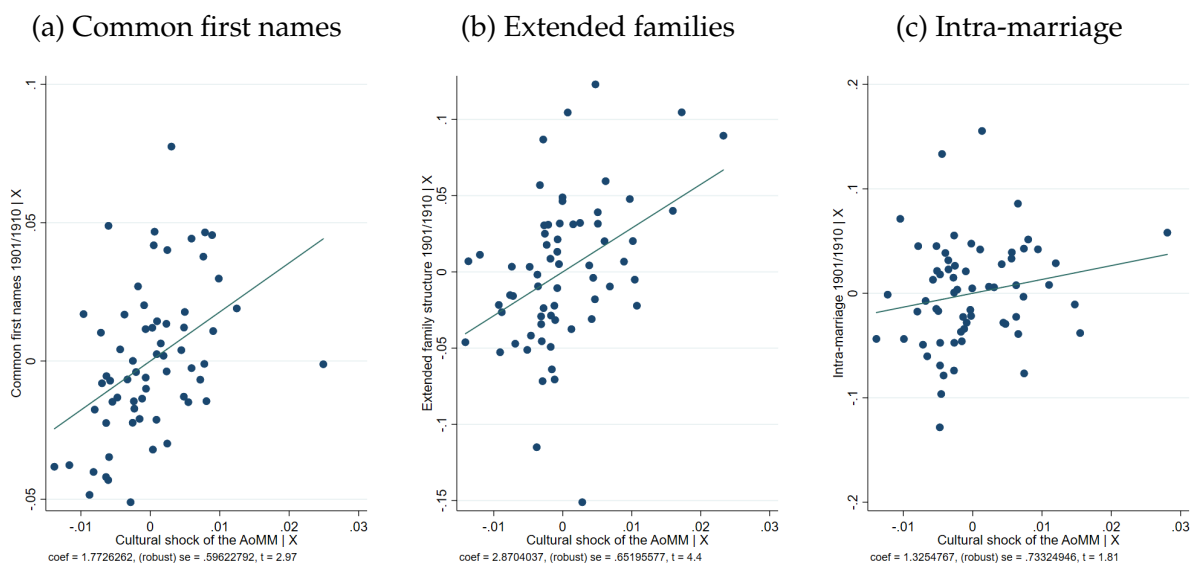
Notes: AV-plots from estimating coefficient β_s in Model 4 across 64 subnational districts. The dependent variable is the prevalence of collectivist norms at the end of the Age of Mass Migration (1901 in DK, 1910 in NO and SE). In Panel A this equals the share of children ($\leq$ 10 years old) with common first names, in Panel B the share of elderly people ($\geq$ 65 years old) who lived with relatives from younger generations, and in Panel C the share of married couples between the ages 20-40 who were born in the same district. The explanatory variable measures the cultural shock of emigration in the preceding decades. Main control variables are the prevalence of collectivist norms and log population size at the beginning of the Age of Mass Migration (1845 in DK, 1865 in NO, 1880 in SE). The *log* number of emigrants in each district is also included as a control along with the children's gender ratio and mean number of first names in Panel A, the share of elderly in the population in Panel B, and the share of people age 20 to 40 who lived outside their district of birth in Panel C. Estimation results can be found in Appendix Table D.19.

tioning of *obedience* and subtracts the mentioning of *independence*. This variable captures contrasting values related to individual autonomy, which is a key difference between collectivism and individualism. To capture more cultural aspects, I construct an additional variable that adds the mentioning of *unselfishness* and subtracts the mentioning of *determination* with a range from −2 to 2.

Second, I make use of questions on the spatial scope of identity. I construct a dummy variable that equals one if the respondent feels a stronger sense of belonging to the city and district where he lives than to the country or world as a whole. This proxies the strength of local identity. Third, I consider the radius of trust. I make use of a set of questions that ask respondents about their level of trust in their family, people in their neighborhood, people they know, people they meet for the first time, and foreigners, respectively. I distinguish between family and everyone else and construct a variable that equals the difference in mean level of trust towards each group.

With this data, I estimate a version of Model 4 that is given by:

$$p_{id}^{post} = \beta_s s_d + \beta_p p_d^{pre} + \beta_X X_{id} + \varepsilon_{id} \, , \tag{5}$$

where p_{id}^{post} is self-reported collectivist beliefs of individual respondent i who lives in Scandinavian district d.[40] p_d^{pre} is the share of children (≤ 10 yrs.) with common first names in the earliest census year for each country and s_d is the sum of cultural shocks between that year and 1920. X_{id} is a set of control variables and ε_d is the error term.

Table 6 reports the results of this estimation. All regressions control for country and survey wave fixed effects, initial population levels, characteristics of the children whose first names proxy initial collectivist norms, and individual respondent characteristics. Standard errors are clustered at the district level. All dependent variables, except the one measuring spatial identity, have been standardized.

Again, two conclusions can be drawn from the results. First, the first row of coefficients indicates that cross-district cultural differences from before the take-off of mass emigration have persisted up until today. Second, the cultural shocks accumulated over the Age of Mass Migration appear to have pushed cultural evolution in a collectivist direction. Respondents who live in districts that experienced the outflow of particularly individualistic people are thus more likely to identify with collectivist attitudes today. These results are robust to controlling for the overall magnitude of emigration in columns (2), (4), (6), and (8).

The results in Table 6 are robust to additional control variables and alternative measures of the cultural shock variable as seen in Appendix Table D.21 and D.22.[41] I can also replicate the analysis with the use of common first names today as the dependent variable in Appendix Table D.23 with the associated AV-plot in Appendix Figure D.2.

5.3 Channels of Persistence

This section has documented the persistence of collectivism, and its shifts in strength due to emigration, in Scandinavia from the mid-19th century until today. The favored explanation of persistence in the context of this paper is that it was driven by *intergenerational cultural transmission*. That is, that persistence was driven by people rather

---

[40]The survey data identifies the NUTS-3 district of each respondent. These are slightly larger than the historical borders used in the rest of the paper. The historical common first names are still defined relative to the historical district, but they now averaged within larger areas.

[41]This includes controlling for the historical use of common last names, geographical characteristics of the district, and the marriage status and education level of the individual respondent.

Table 6. Predicting contemporary collectivist attitudes (WVS/EVS)

Dependent var.	Important child characteristics:				Spatial identity:		Trust in:	
	Obedience vs. independence		+ unselfishness vs. determination		Local vs. national or global		Family vs. others	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Common first names 1845/65/80	0.345***	0.316*	0.457***	0.583***	0.290***	0.332***	0.275**	0.333*
	(0.123)	(0.161)	(0.149)	(0.165)	(0.050)	(0.099)	(0.114)	(0.191)
Cultural shock of the AoMM	5.998**	6.298**	7.007***	5.691***	3.231**	2.857*	7.504***	6.854**
	(2.604)	(2.903)	(1.740)	(1.958)	(1.250)	(1.588)	(2.600)	(3.133)
Baseline controls	Y	Y	Y	Y	Y	Y	Y	Y
Log emigration	N	Y	N	Y	N	Y	N	Y
Mean (sd) of DV	standardized		standardized		0.185 (0.388)		standardized	
No. of districts	50	50	50	50	50	50	50	50
Observations	13,003	13,003	13,003	13,003	10,959	10,959	7,847	7,847
R-squared	0.04	0.04	0.05	0.05	0.07	0.07	0.10	0.10

Notes: The unit of observation is an individual respondent in the pooled WVS and EVS 1996-2018. See the main text for a description of dependent variables. Explanatory variables are measured in the respondents' district of residence (NUTS 3 level): *Common first names* equals the share of children with common first names in 1845 in DK, 1865 in NO, and 1880 in SE. *Cultural shock* is the sum of cultural shocks of emigration from 1845/65/80 until the end of the Age of Mass Migration (AoMM) in 1920. Baseline controls include the *log* population, child gender ratio, and the mean number of first names among children in 1845/65/80. All regressions include country and WVS/EVS wave fixed effects along with individual respondent controls for age, age squared, and gender. Robust standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

than societal structures like formal institutions. If persistence was mainly explained by structures, then the emigration of people with certain cultural traits would carry little predictive power over subsequent cultural norms. The empirical evidence suggests that the opposite was true. When the share of collectivists in the population changed as a result of selective emigration, it was the new, rather than the old, composition of cultural types that persisted over time.[42]

Who was then responsible for the cultural transmission across generations? As highlighted in Section 2.2, the literature on cultural transmission distinguishes between parents and other members of society. The relative importance of these people matters for the strength of the cultural transmission. If transmission only took place from parents, and fertility was constant, then persistence would be perfect. In this case, the cultural shocks associated with the Age of Mass Migration would generate almost perfectly proportional cultural change as most emigrants were young people who would produce or bring their offspring abroad.[43] If other people also mattered, then

---

[42]Moreover, the persistence of the post-emigration composition of cultural types emerged immediately from one decade to the next (as shown in Table 5). This time span is too short for the new composition of cultural types to have impacted the general structures of society.

[43]In the passenger lists, the mean age at emigration was 25 years and 82% were below the age of 35.

persistence would strengthen or weaken depending on the influence of these people. The emigration of particular powerful role models could, for instance, have a wide cultural impact.

Exploring the processes of cultural transmission in depth is beyond the scope of this paper. Appendix Figure D.3 provides evidence that transmission from parents to children was likely significant. Across all parents and children in the census data, having a common first name was associated with a higher likelihood of giving your child a common first name. The correlation, and thus implied transmission, was however not perfect. This suggests an important role for the transmission between non-family members in explaining the patterns of persistence documented in this paper.

6 Conclusion

This paper examines the cultural determinants and consequences of the Age of Mass Migration (1850-1920) in Scandinavia, where around 25 percent of the average population emigrated. The majority of these emigrants settled in North America.

Central to the paper is the hypothesis that people with collectivist cultural traits were less likely to emigrate because they faced higher costs of leaving established social networks behind. To test the hypothesis, I collect data on nearly all stayers and emigrants in Denmark, Norway, and Sweden from historical census records and trans-Atlantic passenger lists. I find that children who grew up in households with conformist naming practices, extended family structures, and strong ties to parents' birthplaces were on average less likely to emigrate later in life. The results are not explained by other relevant childhood circumstances such as socioeconomic status.

Due to selection, mass emigration generated an increase in the stock of collectivist traits in the home population. I approximate the size of this cultural shock with information on the rate and selection of emigration. In a cross-district setting, I find that these shocks significantly predict actual cultural change. This is true throughout the Age of Mass Migration but also up until today. The cultural change that took place in the Age of Mass Migration was thus sufficiently persistent and profound to impact cultural differences in present-day Scandinavia. This suggests that levels of individualism would have been higher had the Age of Mass Migration not occurred.

The potential societal implications of emigration-driven cultural change are of great importance. The period of the Age of Mass Migration was characterized by indus-

trialization and democratization in Scandinavia. Individualism was generally on the rise, in part due to these developments, but it seems conceivable that the collectivist turn caused by emigration played a role in subsequent institutional developments. While economic freedom is high in contemporary Scandinavia, the region is known for its priority of social cohesion and collective insurance. This is particularly clear when contrasted with the liberal capitalism of America, where the majority of the selected emigrants settled. Future research may identify the impact of culture on these developments.

References

[1] Abramitzky, R., Boustan, L. P., and Eriksson, K. (2012), "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration", *American Economic Review*, **102**(5), 1832-56.

[2] Abramitzky, R., Boustan, L. P., and Eriksson, K. (2013), "Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration", *Journal of Development Economics*, **102**(1), 2-14.

[3] Abramitzky, R. and Boustan, L. P. (2017), "Immigration in American Economic History", *Journal of Economic Literature*, **55**(4), 1311-45.

[4] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003), "Fractionalization", *Journal of Economic Growth*, **8**(2): 155-194.

[5] Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2020), "Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States", *Econometrica*, **88**(6), 2329-2368.

[6] Becker, S. O., and Pascali, L. (2019), "Religion, Division of Labor and Conflict: Anti-Semitism in German Regions over 600 Years", *American Economic Review*, **109**(5), 1764-1804.

[7] Bender, H. (2007), "Kilderne på Nettet til Udvandringen fra Danmark", *Working paper.*

[8] Bentzen, J. S. (2019), "Acts of God? Religiosity and Natural Disasters across Subnational World Districts", *Economic Journal*, **129**(622), 2295-2321.

[9] Bhugra, D., and Becker, M. A. (2005), "Migration, Cultural Bereavement and Cultural Identity", *World Psychiatry*, **4**(1), 18-24.

[10] Bisin, A., and Verdier, T. (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *The Quarterly Journal of Economics*, **115**(3), 955-988.

[11] Bisin, A., and Verdier, T. (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences", *Journal of Economic Theory*, **97**(2), 298-319.

[12] Bloothooft, G. and Groot, L. (2008), "Name Clustering on the Basis of Parental Preferences", *Names*, **56**(1), 111-163

[13] Gershman, B. (2017), "Long-Run Development and the New Cultural Economics" in Cervelatti, M. and Sunde, U. (eds.), Demographic Change and Long-Run Development. (Cambridge: MIT Press), 221–261.

[14] Borjas, G. J. (1987), "Self-Selection and the Earnings of Immigrants", *The American Economic Review*, 531-553.

[15] Buggle, J. C. (2020), "Growing Collectivism: Irrigation, Group Conformity and Technological Divergence", *Journal of Economic Growth*, **25**, 147-193.

[16] Burchardi, K. B., Chaney, T., and Hassan, T. A. (2019), "Migrants, Ancestors, and Foreign Investments", *The Review of Economic Studies*, **86**(4), 1448-1486.

[17] Casari, M., De Paola, M., Ichino, A., Marandola, G., Michaeli, M., and Scoppa, V. (2021), "Civicness drain", *Working Paper.*

[18] Cavalli-Sforza, L. L., and Feldman, M. W. (1981), *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton University Press).

[19] Chowdhury, S., Sutter, M., and Zimmermann, K. (2018), "Evaluating Intergenerational Persistence of Economic Preferences: A Large Scale Experiment with Families in Bangladesh", *Working Paper.*

[20] Cohn, Raymond L. (2005), "The Transition from Sail to Steam in Immigration to the United States", *Journal of Economic History*, **65**(2), 469-95.

[21] Daudin, G., Franck, R., and Rapoport, H. (2016), "The Cultural Diffusion of the Fertility Transition: Evidence from Internal Migration in 19th Century France", *Working Paper*.

[22] Davis, L.S. and Abdurazokzoda, F. (2016), "Language, Culture and Institutions: Evidence from a New Linguistic Dataset", *Journal of Comparative Economics*, **44**(3), 541-561.

[23] Docquier, F., and Rapoport, H. (2012), "Globalization, Brain Drain, and Development", *Journal of Economic Literature*, **50**(3), 681-730.

[24] Docquier, F., Tansel, A., and Turati, R. (2020), "Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries", *International Migration Review*, **54**(2), 388-422.

[25] Droller, F. (2017), "Economic Development: Evidence from Settlements in the Pampas", *The Economic Journal*, **128**, 2321-2352.

[26] Eisenbruch, M. (1991), "From Post-Traumatic Stress Disorder to Cultural Bereavement: Diagnosis of Southeast Asian Refugees", *Social Science and Medicine*, **33**(6), 673-680.

[27] Emery, H. B. (2013), "What's In a Name? American Parents' Search for the Perfect Baby Name", *Doctoral dissertation, University of California, Berkeley.*

[28] Enke, B. (2019), "Kinship, Cooperation, and the Evolution of Moral Systems", *The Quarterly Journal of Economics*, **134**(2), 953-1019.

[29] Eriksson, B. (2015), "Dynamic Decades: A Micro Perspective on Late Nineteenth Century Sweden", *Doctoral Dissertation, Lund University.*

[30] Fernández, R., and Fogli, A. (2009), "Culture: An Empirical Investigation of Beliefs, Work, and Fertility", *American Economic Journal: Macroeconomics*, **1**(1), 146-77.

[31] Fogli, A., and Veldkamp, L. (2021), "Germs, Social Networks, and Growth", *The Review of Economic Studies*, **88**(3), 1074-1100.

[32] Fouka, V. (2020), "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I", *The Review of Economic Studies*, **87**(1), 204-239.

[33] Fouka, V., Mazumder, S., and Tabellini, M. (2021), "From immigrants to Americans: Race and Assimilation during the Great Migration", *The Review of Economic Studies*

[34] Fryer, R. G. and Levitt, S. D. (2004), "The Causes and Consequences of Distinctively Black Names", *The Quarterly Journal of Economics*, **119**(3), 767-805.

[35] Giuliano, P., and Tabellini, M. (2020), "The Seeds of Ideology: Historical Immigration and Political Preferences in the United States", *Working Paper*.

[36] Giuliano, P., and Nunn, N. (2021), "Understanding Cultural Persistence and Change", *The Review of Economic Studies*, **88**(4), 1541-1581.

[37] Gorodnichenko, Y., and Roland, G. (2011), "Individualism, Innovation, and Long-Run Growth", *Proceedings of the National Academy of Sciences*, **108**(4), 21316-21319.

[38] Gorodnichenko, Y., and Roland, G. (2011), "Which Dimensions of Culture Matter for Long-Run Growth?", *American Economic Review*, **101**(3), 492-98.

[39] Greif, A. (1994), "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies", *Journal of Political Economy*, **102**(5), 912-950.

[40] Grosjean, P. (2014), "A History of Violence: The Culture of Honor and Homicide in the US South", *Journal of the European Economic Association*, **12**, 1285-1316.

[41] Hatton, T. J., and Williamson, J. G. (1992), *What Drove the Mass Migrations from Europe in the Late Nineteenth Century?*

[42] Heine, S. J., Ruby, M. B. (2010), "Cultural Psychology", *Wiley Interdisciplinary Reviews: Cognitive Science*, **1**(2), 254-266.

[43] Henrich, J. (2020), *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* (Penguin UK).

[44] Hofstede, G. (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations, 2nd ed.* (Sage, Thousand Oaks, CA).

[45] Hornung, E. (2014), "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia", *American Economic Review*, **104**, 84-122.

[46] Indseth, T. (2006), "Forestillinger om Amerika i Norge", *Masteroppgave, University of Oslo*.

[47] International Organization For Migration (IOM). (2019), "World Migration Report", *Report*, Geneva.

[48] Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., and Bonin, H. (2010), "Direct Evidence on Risk Attitudes and Migration", *The Review of Economics and Statistics*, **92**(3), 684-689.

[49] Karadja, M., and Prawitz, E. (2019), "Exit, Voice, and Political Change: Evidence From Swedish Mass Migration to the United States", *Journal of Political Economy*, **127**(4), 1864-1925.

[50] Kashima, E. S.,and Kashima, Y. (1998), "Culture and Language the Case of Cultural Dimensions and Personal Pronoun Use", *Journal of Cross-Cultural Psychology*, **29**(3), 461-486.

[51] Kitayama, S., Ishii, K., Imada, T., Takemura, K., and Ramaswamy, J. (2006), "Voluntary Settlement and the Spirit of Independence: Evidence from Japan's Northern Frontier", *Journal of Personality and Social Psychology,* **91**(3), 369-384.

[52] Lieberson, S. (2000), *A Matter of Taste: How Names, Fashions, and Culture Change* (Yale University Press, New Haven).

[53] Lieberson, S., and Lynn, F. B. (2003), "Popularity as Taste", *Onoma*, **38**, 235-276.

[54] Macfarlane, A. (1978), *The Origins of English Individualism: the Family Property and Social Transition* (Wiley-Blackwell).

[55] Martín Fernández-Sánchez (2020), "Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora", *Working Paper*.

[56] Mateos, P. (2013), *Names, Ethnicity and Populations* (Heidelberg, Germany: Springer).

[57] Meyer-Schwarzenberger, M. (2015), "Indicators of Subject Prominence in 155 National Majority", *Working paper*.

[58] Minnesota Population Center. (2017), *North Atlantic Population Project: Complete Count Microdata Version 2.3.* [Dataset.] (Minneapolis: Minnesota Population Center.)

[59] Munshi, K., and Rosenzweig, M. (2016), "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap", *American Economic Review*, **106**(1), 46-98.

[60] NOS (Norges Offisielle Statistikk) (1921) *VII. 25. Utvandringsstatistikk*. [Emigration statistics.] (Departement for sosiale saker – Kristiania: H. Aschehoug and Co.)

[61] Nunn N. (2012), "Culture and the Historical Process", *Economic History of Developing Regions*, **27**(S1), 108-126.

[62] Olsson, O. and Paik, C. (2016), "Long-Run Cultural Divergence: Evidence from the Neolithic Revolution", *Journal of Development Economics*, **122**, 197-213

[63] Oyserman D., and Lee, S. W. (2008), "Does Culture Influence What and How We Think? Effects of Priming Individualism and Collectivism", *Psychological Bulletin*, **134**, 311-342.

[64] Rapoport, H., Sardoschau, S., and Silve, A. (2020), "Migration and Cultural Change", *Working Paper*.

[65] Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M. and Sobek, M. (2021), *IPUMS USA: Version 11.0.* [Dataset.] (Minneapolis, MN.)

[66] SCB (Statistiska centralbyrån) (1969) *Historical Statistics of Sweden. Part 1. Population 1720-1967.* (Stockholm: AB Allmänna Förlaget.)

[67] Schonberg, W. B., and Murphy, D. M. (1974), "The Relationship Between the Uniqueness of a Given Name and Personality", *The Journal of Social Psychology*, **93**(1), 147-148.

[68] Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., and Henrich, J. (2019), "The Church, Intensive Kinship, and Global Psychological Variation", *Science*, **366**(6466).

[69] Schwartz, S. H. (1994), *Beyond Individualism/Collectivism: New Cultural Dimensions of Values.* (Sage Publications, Inc).

[70] Schwartz, S. H. (2004), "Mapping and Interpreting Cultural Differences Around the World", *International Studies in Sociology and Social Anthropology*, 43-73.

[71] Semmingsen, I., and Haugen, E. I. (1978), *Norway to America: A History of the Migration* (University of Minnesota Press).

[72] Sequeira, S., Nunn, N., and Qian, N. (2020), "Immigrants and the Making of America", *The Review of Economic Studies*, **87**(1), 382-419.

[73] Sønnichsen, O. (2013), *Rejsen til Amerika* (Gyldendal).

[74] Taylor, A. M., and Williamson, J. G. (1997), "Convergence in the Age of Mass Migration", *European Review of Economic History*, **1**(1), 27-63.

[75] Tuccio, M., Wahba, J., and Hamdouch, B. (2019), "International Migration as a Driver of Political and Social Change: Evidence from Morocco", *Journal of Population Economics*, **32**(4), 1171-1203.

[76] Twenge, J. M., Abebe, E. M., and Campbell, W. K. (2010), "Fitting In or Standing Out: Trends in American Parents' Choices for Children's Names, 1880-2007", *Social Psychological and Personality Science*, **1**(1), 19-25.

[77] Uz, I. (2015), "The Index of Cultural Tightness and Looseness among 68 Countries", *Journal of Cross-Cultural Psychology*, **46** (3), 319-335

[78] Van de Vliert, E. (2011), "Climato-Economic Origins of Variation in Ingroup Favoritism", *Journal of Cross-Cultural Psychology*, **42**(3), 494-515.

[79] Vandello, J. A., and Cohen, D. (1999), "Patterns of Individualism and Collectivism across the United States", *Journal of Personality and Social Psychology*, **77**(2), 279.

[80] Varnum, M. E., and Kitayama, S. (2010), "What's in a Name? Popular Names are Less Common on Frontiers", *Psychological Science*, **22**(2), 176-183.

[81] Voigtländer, N., and Voth, H. J. (2012), "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany", *Quarterly Journal of Economics*, **127**(3), 1339-1392.

[82] Zweigenhaft, R. L. (1981), "Unusual Names and Uniqueness", *The Journal of Social Psychology*, **114**(2), 297-298.

A Appendix: Validating the Common Names Proxy

I carry out a number of validation tests to further support the use of common first names as a proxy of collectivism. First, I examine the correlation between the use of common first names and other measures of culture with contemporary data. For this, I collect statistics on the most popular baby names in 2015 across the world. I calculate the prevalence of common first names as the share of newborns with the ten most popular male or female names in the country or district. The result is a dataset with national measures for 38 countries and subnational measures for 13 countries (see Appendix Table D.4 for data sources). Analysis of this data was summarized in the main text in Figure 2, which documented a positive correlation between the use of common first names and survey-based measures of collectivism across and within countries. Here, I present further results.

Table A.1 presents the *cross-country* results. The dependent variable is the widely used 0-100 measure of individualism by Hofstede (2001).[44] Based on surveys of IBM employees, the measure captures a preference for self-reliance and loose social ties. As expected the measure correlates negatively with the use of common first names. The result is robust to relevant controls, including ethnic fractionalization (Alesina et al., 2003). As seen in column (4) it is also robust to controlling for five other cultural measures by Hofstede (2001). The correlation is thus not explained by other key cultural dimensions. National first name patterns also correlate significantly with other more or less known indicators of individualism and collectivism as seen in Appendix Table D.5.[45]

Table A.2 presents the *within-country* results. This table relies on individual responses from the World Values Survey and European Values Study. For the dependent variable, I make use of a set of questions that ask parents about qualities that they regard as important in children. I construct a variable, ranging from −1 to 1, that takes the mentioning of *obedience* and subtracts the mentioning of *independence*. This variable captures contrasting values related to individual autonomy, which is a key difference between collectivism and individualism. To capture more cultural aspects, I construct an additional variable that adds the mentioning of *unselfishness* and subtracts the mentioning of *determination* with a range from −2 to 2. These variables were also used in the analysis on cultural change in Section 5. Country fixed effects ensure that I exploit within-country variation. The results show that individuals who live in districts, where newborns receive common first names, are more likely to value collectivist characteristics.

---

[44]In Figure 2 I subtracted the measure from 100 to capture collectivism, and I restricted the sample to the 23 Western countries for which I have data. Here, all available countries are included.

[45]This holds for survey-based indicators like Schwartz's (1994, 2004) measures of embeddedness, intellectual and affective autonomy, Van de Vliert's (2011) measure of in-group favoritism, and a measure of social tightness from Uz (2015). It also holds when looking at linguistic characteristics like pronoun drop and subject prominence, which have been shown to reflect collectivist and individualistic cultures (Kashima and Kashima, 1998; Abdurazokzoda and Davis, 2014; Meyer-Schwarzenberger, 2015).

Table A.1. Cross-country validation with contemporary data

Dependent var.	Hofstede Individualism Index			
	(1)	(2)	(3)	(4)
Common baby names	-0.465***	-0.444***	-0.466***	-0.367***
	(0.093)	(0.098)	(0.074)	(0.095)
Continent fixed effects	N	Y	Y	Y
Ethnic fractionalization	N	N	Y	Y
Add. Hofstede Controls	N	N	N	Y
Observations	38	38	38	35
R-squared	0.27	0.47	0.76	0.81

Notes: Cross-country correlations between the Hofstede (2001) measure of individualism and the use of common first names, defined as the share of children born in 2015 who were given one of the 10 most popular boys or girls names in their country. All regressions control for *log* population and a dummy for using the Latin alphabet. Additional Hofstede cultural variables include Power Distance, Masculinity, Uncertainty Avoidance, Long Term Orientation, and Indulgence. All variables have been standardized before regression. Robust standard errors are shown in the parentheses. Significance levels are: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Second, I conduct a test with historical data from Sweden. To capture local culture over time, I study the content of 18 local newspapers over the period 1800-1910.[46] I analyze the use of *individualistic vs. collectivist language* based on the frequency of singular (e.g. *I, me, mine*) over plural (e.g. *we, us, our*) pronouns. The use of relatively more singular pronouns reflects that articles, letters, and advertisements were written from an individual perspective, and vice versa from a collective perspective. This is motivated by a broad experimental literature (see review by Oyserman and Lee, 2008).

Results in Table A.3 reveal a significant, negative correlation over time and space between the use of individualistic language and common first names. Exploiting within-newspaper variation, the *log* use of singular pronouns was more prevalent in decades and districts where the use of common first names was less prevalent. This is conditional on the overall size of and use of pronouns in the newspaper, relevant characteristics of the population, newspaper and decade fixed effects. The results hold when considering first person singular pronouns in columns (1) and (2), all singular pronouns in columns (3) and (4), and all singular pronouns that appear next to a verb in columns (5) and (6).

---

[46]The newspapers cover 12 out of 24 districts. Their titles are Tidning för Wenersborgs stad och län (Älvsborg), Blekingeposten (Blekinge), Gotlands tidning (Gotland), Bollnäs tidning (Gävleborg), Dalpilen, Fahlu weckoblad, and Faluposten (Dalarna), Göteborgs weckoblad (Göteborg och Bohus), Gotlands tidning (Gotland), Jönkopingsbladet (Jönköping), Kalmar (Kalmar), Lindesbergs allehanda (Örebro), Norra Skåne (Kristianstad), Östergötlands veckoblad and Östgöta tidning (Östergötland), Wernamo tidning (Jönköping), and Aftonbladet, Post- och Inrikes Tidningar, and Stockholmsposten (Stockholm). They are available for linguistic analysis at *https://spraakbanken.gu.se/*.

Table A.2. Within-country validation with contemporary data

Dependent var.	Important child characteristics:			
	Obedience vs. independence		+ unselfishness vs. determination	
	(1)	(2)	(3)	(4)
Common baby names (district)	0.112***	0.096**	0.085***	0.077**
	(0.036)	(0.039)	(0.032)	(0.034)
Marriage & education (respondent)	N	Y	N	Y
No. of countries	13	12	13	12
No. of districts	199	148	199	148
Observations	50,115	45,406	48,995	44,442
R-squared	0.06	0.08	0.07	0.08

Notes: The unit of observation is an individual surveyed in the World Values Survey and European Values Study. Common baby names are calculated for the subnational district in which the respondents live. All regressions control for country fixed effects, wave fixed effects, *log* number of newborns in the district, and respondent characteristics (age, age squared, and gender). Columns (2) and (4) additionally control for the marriage status and education level of the respondent. All variables have been standardized before regression. Robust standard errors are shown in the parentheses and clustered at the district level. Significance levels are: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3. Validation using Swedish newspaper language, 1800-1910

Dependent var.	Pronoun use (*log* word count)					
	First person singular		All singular		All singular + verb	
	(1)	(2)	(3)	(4)	(5)	(6)
Common first names	-1.264**	-1.432**	-0.163**	-0.185*	-0.173**	-0.217**
	(0.509)	(0.663)	(0.079)	(0.093)	(0.076)	(0.089)
Incl. national newspapers	Y	N	Y	N	Y	N
Newspaper fixed effects	Y	Y	Y	Y	Y	Y
Decade fixed effects	Y	Y	Y	Y	Y	Y
Observations	65	53	65	53	65	53
R-squared	0.93	0.92	1.00	1.00	1.00	1.00

Notes: The sample consists of 18 newspapers observed in two or more decades in the period 1800-1910. 2 newspapers are nationally distributed and excluded in every other column. The dependent variable is the log count of singular pronouns that appear in the newspaper text in a decade. The measure considers first-person singular pronouns in columns (1) and (2), all singular pronouns in columns (3) and (4), and all singular pronouns that appear next to a verb in columns (5) and (6). This is regressed on the share of people with common first names who are born in the same decade and district where the newspaper is published. All regressions include newspaper and decade fixed effects along with newspaper (*log* total word count and *log* word count of all singular and plural pronouns) and population (*log* number, gender ratio, and the mean number of first names of individuals born in the decade and district) controls. All variables have been standardized before regression. Robust standard errors are reported in parentheses with significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Appendix: Construction of Samples and Variables

B.1 Linking Migrants with Census Records

To construct the linked sample, I start with a list of males who towards the end of the Age of Mass Migration had either emigrated or stayed in Scandinavia. Stayers are identified as those still living in the country in the most recent population censuses: The 1901 census in Denmark and the 1910 censuses in Norway and Sweden. Emigrants are identified from the passenger lists up to the year of the most recent census. From this list, I extract a subsample of male emigrants and stayers who were below the age of 15 in any of the earlier censuses.[47] Inspired by related studies (Abramitzky et al., 2013; Eriksson, 2015; Fouka, 2020), I use the following procedure:

1. I construct a sample of potential links based on the string distance between first, middle, and last name, a three-year (±1 year) band around the birth year, and place of origin. String distance is calculated using the Jaro-Winkler algorithm, and I use a score of 0.85 and above as a cutoff to identify potential links.[48] If an individual is recorded without a middle name, he can only be linked to another individual without a middle name, and vice versa. The three-year band around the birth year is allowed because some emigrants were recorded with age rather than birth date. A wider band is not used as people were generally numerate at the time, and there is only very weak evidence of age heaping in the data. The place of origin equals the district of birth for stayers and the country of birth for emigrants, for whom I have no birth district information. These are binding linking criteria. Next, I evaluate the quality of the potential links based on birth year accuracy and, for emigrants, the latest district of residence.

2. I identify as true links those who share the exact birth year and, for emigrants, the same district of residence. Note that stayers already require an exact link on the birth district. I discard cases where multiple links share the same characteristics.

3. I repeat step 2, but now I look for emigrants in the country as a whole. The linking criteria for stayers remain the same. Again, I only keep unique links.

4. I repeat step 2 and 3, letting the birth year vary by one year.

Robustness samples To ensure that the main selection results are not due to a particular aspect of my linking method, I vary the linking requirements and estimate Model 1 with alternative samples. First, I require greater similarity between names across data sources (by evoking higher Jaro-Winkler scores). Second, I only consider links that are unique within a three-year band around the year of birth. This rules out that

---

[47] From 1845 and 1880 in Denmark, 1865 and 1900 in Norway, and 1880, 1890, and 1900 in Sweden. In the Swedish 1890 and 1900 census, I exclude children above the age of 10 to avoid overlaps.

[48] According to tests in Eriksson (2015) this is the optimal score for linking Scandinavian census data.

lower birth year accuracy in the passenger lists matter for the results. Third, I ignore any district information and include links that are unique within the entire country of origin. This ensures that the results are not due to the lack of information on the emigrants' birth districts.

B.2   Constructing the Cultural Shock Variables

The cultural shocks are calculated for each subnational district and decade between 1860 and 1920. They are calculated as the difference between the post- and pre-migration prevalence of people in the population with common first names. The calculation involves three steps:

1. I calculate the share of people with common first names in the population in the *last* year of each decade. Common names are identified, according to the baseline method, relative to the birth district and birth cohort of each individual.

   (a) A challenge is that I only have information on the true residence population in the years that each population census was collected. To cover most of the other decades, I trace the census populations back in time, exclude birth cohorts that would not yet have been born, and add emigrants from the passenger lists who left between the last year of the decade and the census year. For the decades that follow the latest census in each country (1900 and 1910 in Denmark, 1910 in Norway and Sweden), I trace the census population forward in time, subtracting emigrants that left after the census year. This gives me the approximate distribution of cohorts and names in the population in all decades.

   (b) I restrict the population to cohorts that were alive in the last year of the decade and not above the age of 100 in the reference census.

2. I calculate the share of people with common first names in the population in the *first* year of each decade. This is done simply by adding the individual emigrants who left in the decade to the population from Step 1. Information on these emigrants is taken from the passenger lists. They are added to the population in their latest district of residence.

   (a) Since I do not have information on the emigrants' district of birth, I calculate the likelihood that they have a common first name. This equals the share of people in the district with the same first name and birth year whose first name was common in their respective districts of birth. If none in the district shares these characteristics, I calculate the country-wide likelihood or assume that they have an uncommon name.

   (b) Since the passenger lists are not complete, I weigh the emigrants so that their sums equal the true emigration flows across districts and decades. This is

justified by the fact that the passenger lists are representative of the true emigration flows.

3. I subtract the number calculated in Step 1 from that in Step 2. The difference equals the cultural shock.

Robustness measures For robustness, I calculate five alternative versions of the variable. One includes the unweighted emigrants. Another restricts the population to ages below 50 in the final year of the decade. A third excludes emigrants whose first name does not appear in the census. The final two define common first names from their phonetic spelling and within a ±1 year band around the reported year of birth.

C   Appendix: Selection in Migration, Group-Level Results

As discussed in Section 4.1, the individual-level results may be biased by the fact that I cannot link every individual to his childhood household. Although I use different strategies to minimize this bias, concerns still remain. In this section, I therefore carry out a group-level analysis with *all* emigrants and stayers. This includes females, who were left out of the individual-level analysis.

The unit of observation is a group of individuals who share the same first name, gender, birth year, and district of residence. The baseline sample consists of groups that were below the age of 15 in any population census. For each group, I calculate their propensity to emigrate by dividing the number of emigrants in the passenger lists with the original census population.[49] I also calculate the average of all childhood characteristics in the census, including collectivist norms. For example, the common first name indicator now equals the share of the group's individuals whose first name was common in their year and district of birth.

Compared with the individual-level data, the group-level variables are less precisely measured. First, they are averages over individuals. Second, emigrants who have moved to another district since the age of 15 will be mis-grouped relative to their younger selves in the census. Nevertheless, the data allows for an almost fully representative analysis of selection in migration.

With the group-level data, I estimate the following model:

$$m_{nsbdc} = \beta_p p_{nsbdc} + \phi_d + \phi_y + \phi_c + \beta_X X_{nsbdc} + \varepsilon_{nsbdc}, \tag{6}$$

where m_{nsbdc} is the total emigration rate (1868-1920) in the group of individuals who share the same first name n, gender s, birth year b, district of residence d, and who were below the age of 15 in census c. p_{nsbdc} is a proxy of collectivist norms in the childhood household and X_{nsbdc} is a set of childhood control variables. These are largely the same as in Model 1, but now they are averaged over the individuals in the group. ϕ_d, ϕ_y, and ϕ_c denote residence district, birth decade, and census fixed effects.

Results are reported in Table C.1. Standard errors are clustered at the residence district $\times$ birth decade $\times$ census level. Columns (1) to (6) include groups for which I have measures of all explanatory and control variables. The coefficients on all proxies are negative and robust to including relevant childhood controls. This implies that groups that were more likely to have grown up in collectivist households were less likely to emigrate later in life.

The results in columns (1) to (6) should be seen as conservative, because they exclude 6.4 percent of the emigrants who do not share group characteristics with anyone in the census. This is more likely to happen for emigrants whose first name is uncommon. Treating them as stayers creates a downward bias on the coefficients. In column (7)

---

[49]If a cohort is observed in more than one census, I use the earliest census to calculate the childhood variables. The census population includes emigrants that left before the census.

I include these emigrants and assume that they had uncommon first names and that their propensity to emigrate was 100 percent. The associated coefficient is considerably larger than in columns (1) and (2). It implies that groups, where all (compared to no) individuals were given a common first name, were 1.2 percentage points more likely to emigrate.

Table C.1. Collectivist norms and emigration: Group-level results

Included emigrants	In groups that are observed in the census						All
Dependent var.	Rate of emigration (0-100)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Common first name	-0.413***	-0.644***					-1.171***
	(0.105)	(0.138)					(0.121)
Extended family			-2.230***	-2.053***			
			(0.381)	(0.377)			
Parents same birthplace					-1.023***	-0.770***	
					(0.212)	(0.198)	
Childhood controls	N	Y	N	Y	N	Y	N
Mean of DV	8.83	8.83	8.83	8.83	8.86	8.86	9.48
SD of DV	14.40	14.40	14.40	14.40	14.02	14.02	16.47
Observations	1,079,778	1,079,778	1,079,778	1,079,778	949,281	949,281	1,152,127
R-squared	0.24	0.25	0.24	0.25	0.26	0.26	0.21

Notes: The unit of observation is a group of individuals who share the same first name, gender, birth year, and district of residence, observed in the earliest available census. The propensity to emigrate equals the number of emigrants in the passenger lists divided by the original census population. Explanatory and control variables are averages over individuals in the group. Columns (1)-(6) include emigrants whose groups are observed in the census. Column (7) includes all emigrants and assumes a common name indicator of zero and an emigration rate of 100. Each observation is weighted by its size in the census population. The sum of weights equals 8.12-8.27 mil. individuals in columns (1)-(6) and 8.35 mil. in column (7). All regressions include census, birth decade, and residence district fixed effects along with controls for age in the census and number of first names. Childhood controls include gender, log family size, being firstborn, log number of siblings, log birth order, having a foreign-born parent, having moved since birth, having a common last name, having a first name shared by an older family member, urban residence, having servants employed, being related to the household head, and having a parent on poor relief. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results are robust to defining groups differently and including groups that were above the age of 15 in a given census. Specifically, Appendix Table D.11 shows that the results are robust to defining groups by the phonetic spelling of first names, and to calculating the census variables as three-year moving average over birth years. This is to ensure that discrepancies in reported names and birth years do not drive the results. The results are not exclusive to groups that were below the age of 15 in the censuses. Results in Appendix Table D.12 include all birth cohorts, using the common first name indicator as a proxy and excluding childhood controls. The results indicate that selection in migration was a broad and significant phenomena in the Scandinavian populations.

I can also assess *how* common a first name has to be in order to be associated with lower emigration rates. Figure C.1 shows that defining more names as common

weakens the relationship with emigration. Once the 20 most common names are included, there is no significant difference in emigration rates. The 20 most common names are carried by 62 percent of the sample.

Figure C.1. Emigration and the commonness of first names
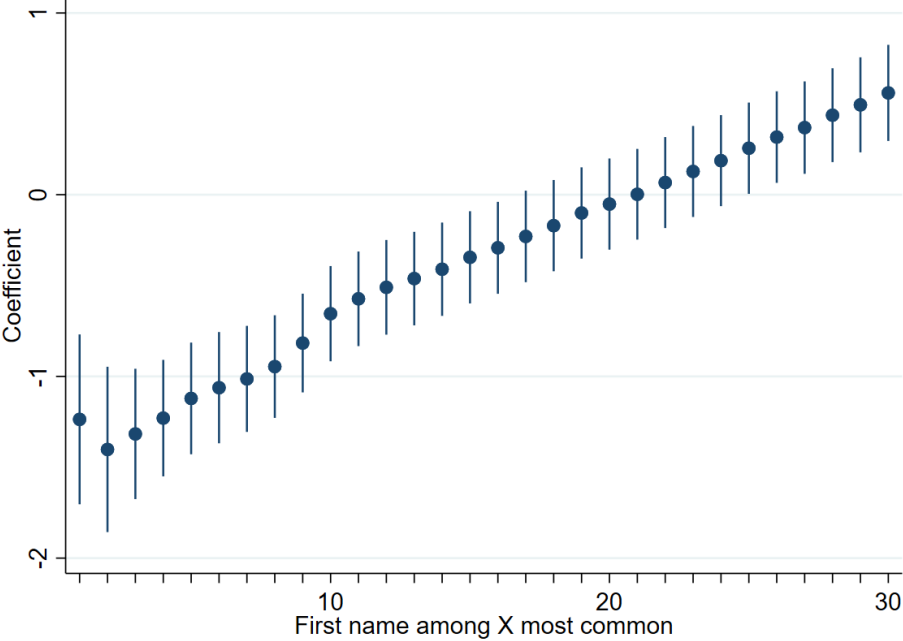
Notes: Coefficients from estimating the full model with common first names from Column (2) in Table C.1, but instead of looking at the ten most common first names, I vary this from 1-30. The plot includes 95% confidence intervals.

D Appendix: Additional Figures and Tables

Figure D.1. Total rate and cultural shock of emigration 1860-1920
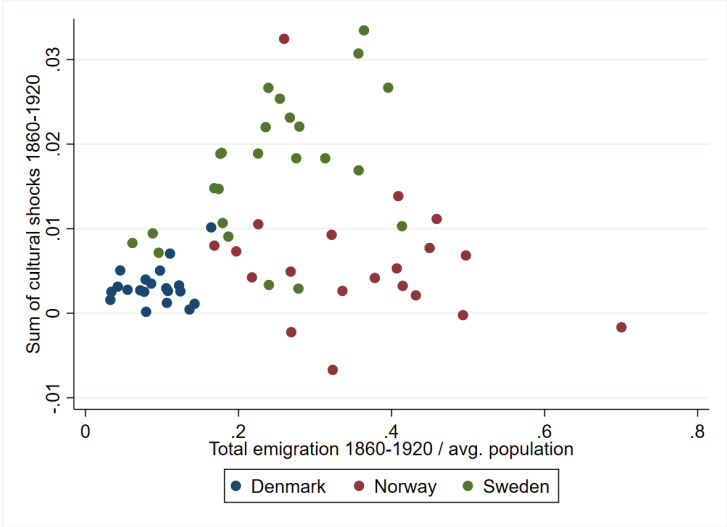
Notes: Scatter plot across 64 Scandinavian districts. The x-axis shows the total rate of emigration measured as the total number of emigrants 1860-1920 divided by the average population in the same period. The y-axis shows the sum of cultural shocks of emigration over the decades 1806-1920, which measures the overall increase in the share of people with common first names in the home population due to selective emigration.

Figure D.2. Persistence of cultural shocks in the long run, AV-plot

coef = 1.6606156, (robust) se = .49801364, t = 3.33
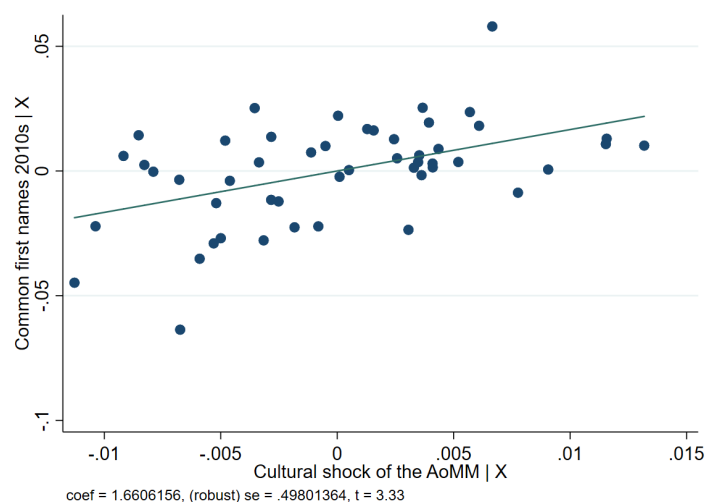
Notes: AV-plot of the cross-district relationship between contemporary naming patterns and the sum of cultural shocks over the Age of Mass Migration. The dependent variable is the share of newborns born in the 2010s with common first names. Main control variables are the prevalence of common first names and log population size at the beginning of the Age of Mass Migration (1845 in DK, 1865 in NO, 1880 in SE). The *log* number of emigrants in each district is also included as a control. Estimation results can be found in Appendix Table D.23.

Figure D.3. Parent-child correlation in common first names
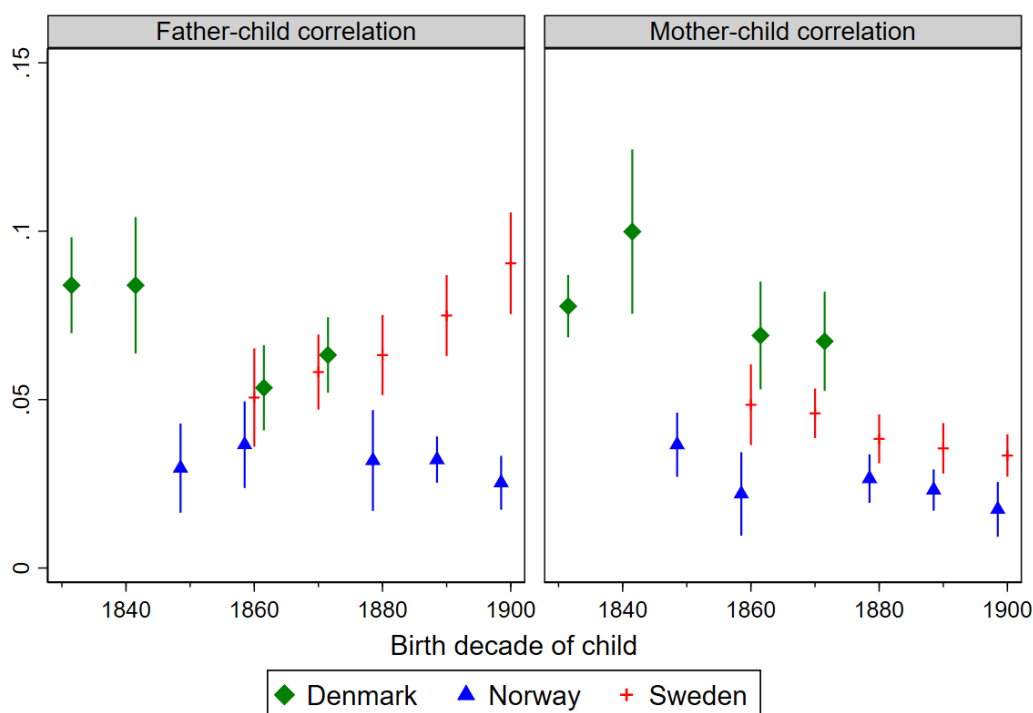
Notes: Coefficients from regressing the common first name indicator of the firstborn child on that of both parents. The 95% confidence intervals are illustrated as vertical lines around the coefficients. The father and mother enter in the same regression. One regression is run separately for each group of children who are born in the same decade and the same country. All regressions include fixed effects for the birth district of the child and birth decades of the father and mother. Control variables include the child's gender, age, age squared, and the age of mother and father.

Table D.1. Cross-validation of emigration data 1860-1920 (Norway and Sweden)

Dependent var.	No. of emigrants (national accounts)		
	(1)	(2)	(3)
No. of emigrants (passenger lists)	1.013***	1.008***	0.787***
	(0.051)	(0.052)	(0.071)
Decade fixed effects	Y	Y	Y
Country fixed effects	N	Y	Y
District fixed effects	N	N	Y
Mean (sd) of DV	6936 (5434)	6937 (5434)	6938 (5434)
Observations	264	264	264
R-squared	0.69	0.69	0.80

Notes: OLS regressions comparing the absolute number of emigrants recorded each decade 1860-1910 across 44 Swedish and Norwegian subnational districts. Robust standard errors in parentheses with significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table D.2. Emigration and cross-census demographic change

Dependent var.	Change in population census					
	(1)	(2)	(3)	(4)	(5)	(6)
Emigrants	-0.749***	-1.413***	-0.720***	-1.397***	-0.772***	-1.414***
	(0.030)	(0.028)	(0.030)	(0.028)	(0.030)	(0.028)
Log census number	N	Y	N	Y	N	Y
Census fixed effects	N	N	Y	Y	Y	Y
Demography fixed effects	N	N	N	N	Y	Y
Mean of DV	0.452	0.452	0.452	0.452	0.452	0.452
SD of DV	10.22	10.22	10.22	10.22	10.22	10.22
Observations	3,236,472	3,236,472	3,236,472	3,236,472	3,236,472	3,236,472
R-squared	0.01	0.03	0.06	0.06	0.08	0.10

Notes: Cross-validation of emigration data, OLS regressions. The dependent variable is the absolute change from one population census to the next in the number of people who are born in the same year, reside in the same district, and carry the same first name. The explanatory variable is the number of emigrants with the same demographic characteristics who emigrated between the respective census years. The sample is restricted to birth cohorts that would have been below the age of 80 in the later census year. *Demography fixed effects* include gender, birth year, and district fixed effects. Robust standard errors in parentheses with significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table D.3. Most frequent first names among children in select censuses

Census:	Denmark 1901		Norway 1910		Sweden 1910	
Rank	Boys	Girls	Boys	Girls	Boys	Girls
1	Jens (7%)	Anna (6%)	Ole (2%)	Anna (3%)	Karl (11%)	Anna (7%)
2	Hans (6%)	Ane (4%)	Johan (2%)	Astrid (1%)	Erik (6%)	Elsa (3%)
3	Niels (6%)	Karen (4%)	Karl (2%)	Ingeborg (1%)	Nils (5%)	Karin (3%)
4	Karl (3%)	Marie (4%)	Hans (2%)	Gudrun (1%)	Johan (4%)	Marta (3%)
5	Peter (2%)	Johanne (3%)	Olav (1%)	Borghild (1%)	Gustaf (4%)	Astrid (2%)
6	Johannes (2%)	Ellen (2%)	Olaf (1%)	Margit (1%)	Sven (4%)	Signe (2%)
7	Carl (2%)	Agnes (2%)	Einar (1%)	Marie (1%)	John (2%)	Ester (2%)
8	Anders (2%)	Maren (2%)	Arne (1%)	Olga (1%)	Axel (2%)	Rut (2%)
9	Kristian (2%)	Ingeborg (1%)	Harald (1%)	Jenny (1%)	Bror (2%)	Svea (2%)
10	Christian (2%)	Olga (1%)	Nils (1%)	Marta (1%)	Oskar (2%)	Edit (1%)

Notes: The most frequent first names among girls and boys aged 0-10 in the most recent census from each country. The shares presented in the parentheses are calculated relative to all girls and boys separately.

Table D.4. Validation using contemporary indicators: Countries and sources

Continent	Country	Country	District	Source
Africa	South Africa	x		Imbizo Centre
Americas	Argentina	x		Registro Civil de la Ciudad de Buenos Aires (Only Buenos Aires)
Americas	Canada	x	x	Governments of Alberta, British Columbia, Manitoba, Nova Scotia, Ontario, Quebec, Saskatchewan, and Northern Territories
Americas	Chile	x		Chile Registro Civil
Americas	Colombia	x		Registraduria Nacional del Estado Civil
Americas	Peru	x		Registro Nacional de Identificacion y Estado Civil (Only 3 first months of 2015)
Americas	United States	x	x	US Social Security
Asia	Israel	x		Central Bureau of Statistics Israel (Only Jewish population)
Asia	Philippines	x		Philippine Statistics Authority
Europe	Austria	x	x	Statistik Austria
Europe	Belgium	x	x	Statistics Belgium
Europe	Bulgaria	x		National Statistical Institute (NSI)
Europe	Czech Republic	x		Ministry of the Interior of the Czech Republic
Europe	Denmark	x	x	Danmarks Statistik
Europe	Estonia	x		Estonia Ministry of the Interior
Europe	Finland	x		Population Register Centre
Europe	France	x	x	Institut national de la statistique
Europe	Germany	x		blog.beliebte-vornamen.de (independent collection of city data)
Europe	Hungary	x		Belugyminiszterium Nyilvantartasok Vezeteseert Felelos Helyettes Allamtitkarsag
Europe	Iceland	x		Statistics Iceland
Europe	Ireland	x		The Central Statistics Office of Ireland (SCO)
Europe	Italy	x	x	Istat - Istituto Nazionale di Statistica
Europe	Latvia	x		Office of Citizenship and Migration Affairs
Europe	Lithuania	x		Population Register Service
Europe	Malta	x		Malta National Statistics Office (Phonetic spellings combined)
Europe	Moldova	x		Moldova Civil Status Service
Europe	Netherlands	x		Meertens Institut
Europe	Norway	x	x	StatBank Norway
Europe	Poland	x		Ministry of Interior and Administration
Europe	Portugal	x		Instituto dos Registos e do Notariado
Europe	Romania	x		Ministry of Administration and Interior
Europe	Slovakia	x		Interior Ministry
Europe	Slovenia	x		Republic of Slovenia Statistical Office RS
Europe	Spain	x	x	Instituto National de Estadistica
Europe	Sweden	x	x	Statistics Sweden
Europe	Switzerland	x	x	Federal Statistical Office Switzerland
Europe	United Kingdom	x	x	Office for National Statistics
Oceania	Australia	x	x	Registers of Births, Deaths and Marriages (from NSW, Victoria, Queensland, South Australia, Western Australia, and Northern Territory)

Notes: This table lists the countries for which contemporary first names data was collected. The sources for the collection are also noted. The sources are found by searching the web pages of national statistical offices, newspapers, and broader collections of first names databases like behindthename.com and nancy.cc. All searches are done in English and the national language (using Google Translate).

Table D.5. Cross-country validation of the common first name proxy: Additional indicators

Dependent var.	Survey Based Measures					Linguistic Characteristics		
	Embeddedness (1)	Affective Autonomy (2)	Intellectual Autonomy (3)	In-Group Favouritism (4)	Social Tightness (5)	Pronoun Drop (KK 1998) (6)	Pronoun Drop (AD 2014) (7)	Subject Prominence (8)
Common baby names	0.550** (0.215)	-0.728*** (0.223)	-0.342 (0.226)	0.496*** (0.148)	0.467** (0.172)	0.501*** (0.131)	0.666** (0.299)	-0.378** (0.161)
Observations	28	28	28	39	32	32	37	33
R-squared	0.37	0.44	0.34	0.45	0.28	0.66	0.45	0.43

Notes: Additional country-level validation regressions. Embeddedness, Intellectual and Affective Autonomy are from Schwartz (1994, 2004), In-Group Favouritism from Van de Vliert (2011), and Social Tightness from Uz (2015). The original Pronoun Drop dummy is from Kashima and Kashima (1998) and the extended from Abdurazokzoda and Davis (2014). The Subject Prominence indicator of individualism is from Meyer-Schwarzenberger (2015). Common baby names measures the share of children born in 2015 who were given one of the 10 most popular boys or girls names in their country. All regressions control for continent fixed effects, log population, and a dummy for using the latin alphabet. All variables have been standardized before regression. Robust standard errors are shown in the parentheses. Significance levels are: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.6. Collectivism and emigration: Alternative linking methods

Dependent var.	Emigration			
	(1)	(2)	(3)	(4)
Panel A				
Common first name	-0.0157***	-0.0179***	-0.0093***	-0.0045***
	(0.002)	(0.002)	(0.001)	(0.002)
Mean (sd) of DV	0.106 (0.307)	0.095 (0.293)	0.085 (0.279)	0.084 (0.277)
Observations	699,315	709,323	480,889	323,590
R squared	0.15	0.14	0.16	0.13
Panel B				
Extended family	-0.0072***	-0.0061***	-0.0052**	-0.0036*
	(0.002)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.106 (0.307)	0.095 (0.293)	0.085 (0.279)	0.084 (0.277)
Observations	699,315	709,323	480,889	323,590
R-squared	0.15	0.14	0.16	0.13
Panel C				
Parents intra-marriage	-0.0087***	-0.0081***	-0.0071***	-0.0038***
	(0.001)	(0.001)	(0.001)	(0.001)
Mean (sd) of DV	0.104 (0.305)	0.093 (0.290)	0.083 (0.275)	0.081 (0.273)
Observations	624,862	633,064	427,822	285,198
R-squared	0.15	0.14	0.15	0.13
Linking method	Jaro-Winkler cutoff at 0.9	Jaro-Winkler cutoff at 0.95	Unique within birth year band	Disregard district information

Notes: This table replicates the full specification from column (4) in Table 3, using alternative linking methods (described in Appendix B.1). The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses, who are linked to either the passenger lists (emigration = 1) or the most recent census (emigration = 0) for each country. Explanatory variables proxy collectivist norms in the childhood household. All regressions include census, birth decade, and birth district fixed effects. Controls include demographics of the family, past migration, socioeconomic status, and additional names characteristics (detailed in Table 3). For columns (1) and (2) I use more conservative Jaro-Winkler cutoffs that require links to have a stronger string similarity between first, middle, and last names. For column (3) I only consider links that are unique within a three-year band around the birth year. For column (4) I ignore any district information and include links that are unique within the entire country of origin. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.7. Collectivism and emigration: Probit estimation

Dependent var.	Emigration			
	(1)	(2)	(3)	(4)
Panel A				
Common first name	-0.0847***	-0.0741***	-0.0645***	-0.0606***
	(0.009)	(0.009)	(0.009)	(0.009)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)
Observations	691,969	691,969	691,969	691,969
Panel B				
Extended family	-0.0579***	-0.0688***	-0.0678***	-0.0528***
	(0.013)	(0.014)	(0.014)	(0.013)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)
Observations	691,969	691,969	691,969	691,969
Panel C				
Parents intra-marriage	-0.0849***	-0.0736***	-0.0757***	-0.0600***
	(0.010)	(0.010)	(0.010)	(0.008)
Mean (sd) of DV	0.108 (0.310)	0.108 (0.310)	0.108 (0.310)	0.108 (0.310)
Observations	618,934	618,934	618,934	618,934
Demography and past migration	N	Y	Y	Y
Additional name characteristics	N	N	Y	Y
Socioeconomic status	N	N	N	Y

Notes: This table replicates the results from Table 3, using probit rather than OLS estimation. The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses, who are linked to either the passenger lists (emigration = 1) or the most recent census (emigration = 0) for each country. Explanatory variables proxy collectivist norms in the childhood household. All regressions include census, birth decade, and birth district fixed effects. Control variables are detailed in Table 3. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.8. Collectivism and emigration: Additional fixed effects

Dependent var.	Emigration				
	(1)	(2)	(3)	(4)	(5)
Panel A					
Common first name	-0.0118***	-0.0075***	-0.0137***	-0.0100***	-0.0115***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.109 (0.312)	0.102 (0.303)	0.110 (0.313)
Observations	691,682	691,290	687,953	599,008	691,961
R squared	0.15	0.16	0.17	0.15	0.15
Panel B					
Extended family	-0.0088***	-0.0071***	-0.0091***	-0.0066***	-0.0079***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.109 (0.312)	0.102 (0.303)	0.110 (0.313)
Observations	691,682	691,290	687,953	599,008	691,961
R-squared	0.15	0.16	0.17	0.15	0.15
Panel C					
Parents intra-marriage	-0.0090***	-0.0083***	-0.0102***	-0.0071***	-0.0098***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean (sd) of DV	0.108 (0.310)	0.107 (0.310)	0.107 (0.309)	0.101 (0.301)	0.108 (0.310)
Observations	618,686	618,267	615,470	568,788	618,933
R-squared	0.15	0.16	0.17	0.15	0.14
Additional fixed effects	Birth municipality	First name	Last name	Father's occupation	Birth order

Notes: This table replicates the full specification from column (4) in Table 3, including additional fixed effects. The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses, who are linked to either the passenger lists (emigration = 1) or the most recent census (emigration = 0) for each country. Explanatory variables proxy collectivist norms in the childhood household. All regressions include census, birth decade, and birth district fixed effects. Controls include demographics of the family, past migration, socioeconomic status, and additional names characteristics (detailed in Table 3). Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.9. Collectivism and emigration: Additional robustness tests

Dependent var.	Emigration				Out-migration
	(1)	(2)	(3)	(4)	(5)
Panel A					
Common first name	-0.0126***	-0.0029***	-0.0098***	-0.0140***	-0.0072***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Mean (sd) of DV	0.110 (0.313)	0.017 (0.129)	0.096 (0.295)	0.123 (0.328)	0.120 (0.325)
Observations	691,969	626,564	681,354	618,060	615,949
R squared	0.14	0.02	0.14	0.19	0.30
Panel B					
Extended family	-0.0092***	-0.0024***	-0.0064***	-0.0103***	-0.0136***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)
Mean (sd) of DV	0.110 (0.313)	0.017 (0.129)	0.096 (0.295)	0.123 (0.328)	0.120 (0.325)
Observations	691,969	626,564	681,354	618,060	615,949
R-squared	0.14	0.02	0.14	0.19	0.30
Panel C					
Parents intra-marriage	-0.0061***	-0.0025***	-0.0083***	-0.0172***	-0.0389***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)
Mean (sd) of DV	0.108 (0.310)	0.016 (0.128)	0.094 (0.292)	0.120 (0.325)	0.116 (0.320)
Observations	618,934	561,704	609,615	554,818	552,385
R-squared	0.14	0.02	0.14	0.18	0.29
Robustness check	Excl. district fixed effects	Excl. adult emigrants	Excl. child emigrants	Excl. out-migrants	Excl. emigrants

Notes: This table tests the robustness of the full specification results from column (4) in Table 3. The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses, who are linked to either the passenger lists (emigration = 1) or the most recent census (emigration = 0) for each country. Explanatory variables proxy collectivist norms in the childhood household. All regressions include census, birth decade, and birth district fixed effects. Controls include demographics of the family, past migration, socioeconomic status, and additional names characteristics (detailed in Table 3). Column (1) excludes birth district fixed effects to allow for variation across districts. Columns (2) and (3) exclude emigrants who emigrated above or below the age of 15, respectively. Columns (4) and (5) analyze the decision to emigrate and out-migrate (from one's birth district to another) separately. To ease the comparison of the coefficients, the alternative migration group is excluded in each of these regressions. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.10. Collectivism and emigration: Alternative measures of common first names

Dependent var.	Emigration								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Common first name	-0.0114*** (0.002)	-0.0067*** (0.001)	-0.0119*** (0.002)	-0.0113*** (0.002)	-0.0113*** (0.002)	-0.0119*** (0.002)	-0.0088*** (0.001)	-0.0054*** (0.001)	-0.0073*** (0.001)
First names	Phonetic spelling	Initial (first letter)	Abbreviation (first three letters)	Officially recognized names	Non-rare names	Non-emigration adjusted	Country level	Oldest sibling	Log share w. same name (std)
Mean (sd) of DV	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.111 (0.314)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)	0.110 (0.313)
Observations	691,969	691,969	691,969	653,285	690,204	691,969	691,969	688,955	691,969
R squared	0.15	0.15	0.15	0.14	0.15	0.15	0.15	0.15	0.15

Notes: This table replicates the full specification from Panel A, column (4) in Table 3, using alternative measures of common first names. The sample consists of male children ($\leq$ 15 yrs.) in the earlier censuses, who are linked to either the passenger lists (emigration = 1) or the most recent census (emigration = 0) for each country. All regressions include census, birth decade, and birth district fixed effects. Controls include demographics of the family, past migration, socioeconomic status, and additional names characteristics (detailed in Table 3). In columns (1)-(5) I alter the baseline commonness indicator by using different spellings or restricting the sample of included first names. Officially recognized names are names that contemporary authorities have approved for use in each country. Non-rare names are shared by more than 10 people in the census. Column (6) uses a commonness indicator that is not adjusted for pre-census emigration. Column (7) uses an indicator calculated at the country rather than district of birth level. Column (8) considers the common first name indicator of the oldest brother in the household. In column (9) commonness is measured as the (standardized) log share of the past ten cohorts in the birth district that share the same name. Standard errors are clustered at the census $\times$ birth decade $\times$ district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.11. Collectivist norms and emigration, group-level results: Robustness

Included emigrants	All	In groups that are observed in the census					
Collectivist norm	Common first name			Extended family		Parents same birthpl.	
Dependent var.	Rate of emigration (0-100)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Group link based on phonetic spelling of first names							
Collectivist norm	-0.951***	-0.468***	-0.620***	-2.174***	-2.186***	-0.631***	-0.649***
	(0.100)	(0.089)	(0.125)	(0.324)	(0.335)	(0.189)	(0.188)
Mean of DV	9.50	9.50	9.12	9.12	9.12	9.14	9.14
SD of DV	14.50	14.50	13.15	13.15	13.15	12.83	12.83
Observations	896,063	850,830	850,830	850,830	850,830	761,736	761,736
R-squared	0.27	0.31	0.31	0.31	0.31	0.32	0.33
Panel B: Group link based on smoothed 3-year average							
Collectivist norm	-1.425***	-0.960***	-1.049***	-3.767***	-3.351***	-2.392***	-1.494***
	(0.133)	(0.118)	(0.157)	(0.658)	(0.645)	(0.326)	(0.328)
Mean of DV	9.91	9.48	9.48	9.48	9.48	9.50	9.50
SD of DV	19.54	18.41	18.41	18.41	18.41	17.80	17.80
Observations	1,895,301	1,828,578	1,828,578	1,828,578	1,828,578	1,602,775	1,602,775
R-squared	0.17	0.17	0.18	0.17	0.18	0.19	0.19
Childhood controls	N	N	Y	N	Y	N	Y

Notes: The unit of observation is a group of individuals who share the same first name, gender, birth year, and district of residence, and who were below the age of 15 in the censuses. Variables equal averages over the individuals in the group. Column (1) includes all emigrants and assumes a common name indicator of zero and an emigration rate of 100. Columns (2)-(7) include emigrants whose groups are observed in the census. Each observation is weighted by its size in the census population. All regressions include census, birth decade, residence district fixed effects, and controls for age in the census and number of first names. Childhood controls are listed in Table C.1. Standard errors are clustered at the census × birth decade × district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.12. Collectivist norms and emigration, group-level results: All cohorts

Group link based on	Earliest available census		Phonetic spelling of names		Smoothed 3-year average	
Dependent var.	Rate of emigration					
	(1)	(2)	(3)	(4)	(5)	(6)
Common first name	-1.048***	-0.432***	-0.743***	-0.330***	-1.415***	-0.913***
	(0.102)	(0.091)	(0.078)	(0.072)	(0.138)	(0.123)
Including all emigrants	Y	N	Y	N	Y	N
Mean of DV	9.807	9.270	9.829	9.510	10.34	9.861
SD of DV	17.38	15.98	15.29	14.35	20.93	19.93
Observations	2,333,138	2,252,991	1,804,699	1,757,128	1,941,900	1,903,089
R squared	0.22	0.24	0.28	0.31	0.17	0.17

Notes: The unit of observation is a group of individuals who share the same first name, gender, birth year, and district of residence. Variables equal averages over the individuals in the group. Columns (1), (3), (5), and (7) include all emigrants and assumes a common name indicator of zero and an emigration rate of 100. The other columns include only emigrants whose groups are observed in the census. Each observation is weighted by its size in the census population. All regressions include census, birth decade, residence district fixed effects, and controls for age in the census and number of first names. Standard errors are clustered at the census × birth decade × district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.13. Collectivist norms and emigration style

Dependent var.	Emigrate with family		Emigrate with townspeople		Emigrate on joint ticket	
	(1)	(2)	(3)	(4)	(5)	(6)
Common first name	0.0081***	0.0082***	0.0099***	0.0096***	0.0075***	0.0091***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Common last name	N	Y	N	Y	N	Y
Mean (sd) of DV	0.39 (0.49)	0.39 (0.49)	0.70 (0.46)	0.70 (0.46)	0.35 (0.48)	0.35 (0.48)
Observations	1,462,134	1,462,134	1,411,676	1,411,676	1,055,797	1,055,797
R-squared	0.22	0.22	0.08	0.08	0.26	0.26

Notes: OLS regressions. The unit of observation is an emigrant observed in the passenger lists. Commonness of first (and last) names is measured in the earliest available census relative to the individual's birth cohort, gender, and last district of residence. Regressions include birth decade, birth district, emigration decade, and census fixed effects along with controls for age and age squared at emigration, gender, and number of first names. Standard errors are double clustered at the census $\times$ birth decade $\times$ district and family level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.14. Collectivism and emigration style: Robustness

Dependent var.	Emigrate with family		Emigrate with townspeople		Emigrate on joint ticket	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Common names based on phonetic spelling						
Common first name	0.0062***	0.0063***	0.0098***	0.0095***	0.0072***	0.0088***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean (sd) of dep. var.	0.39 (0.49)	0.39 (0.49)	0.70 (0.46)	0.70 (0.46)	0.35 (0.48)	0.35 (0.48)
Observations	1,462,128	1,462,128	1,411,670	1,411,670	1,055,791	1,055,791
R-squared	0.22	0.22	0.08	0.08	0.26	0.26
Panel B: Common names within a ±1 birth year band						
Common first name	0.0085***	0.0086***	0.0107***	0.0104***	0.0082***	0.0098***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean (sd) of dep. var.	0.39 (0.49)	0.39 (0.49)	0.70 (0.46)	0.70 (0.46)	0.35 (0.48)	0.35 (0.48)
Observations	1,462,134	1,462,134	1,411,676	1,411,676	1,055,797	1,055,797
R-squared	0.22	0.22	0.08	0.08	0.26	0.26
Panel C: Common names from census where cohort has adult age (max 40 yrs.)						
Common first name	0.0082***	0.0087***	0.0107***	0.0106***	0.0082***	0.0103***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean (sd) of DV	0.39 (0.49)	0.39 (0.49)	0.70 (0.46)	0.70 (0.46)	0.35 (0.48)	0.35 (0.48)
Observations	1,462,134	1,462,134	1,411,676	1,411,676	1,055,797	1,055,797
R-squared	0.22	0.22	0.08	0.08	0.25	0.26
Common last name	N	Y	N	Y	N	Y

Notes: OLS regressions. The unit of observation is an emigrant observed in the passenger lists. Commonness of first (and last) names is measured in the earliest available census relative to the individual's birth cohort, gender, and district of residence. Panel A uses the phonetic spelling of first names, Panel B the census in which the emigrant would be closest to (but maximum) 40 years old, and Panel C defines common first names within ±1 year band around the reported year of birth. Regressions include birth decade, birth district, emigration decade, and census fixed effects along with controls for age and age squared at emigration, gender, and number of first names. Standard errors are double clustered at the census $\times$ birth decade $\times$ district and family level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.15. Settlement of Scandinavian migrants in the United States

Dependent var.	People of own nationality in county (*log*)		Spouse same nationality		Scandinavian name index of oldest US-born child	
	(1)	(2)	(3)	(4)	(5)	(6)
Common first name	0.179***	0.179***	0.044***	0.043***	0.323***	0.323***
	(0.028)	(0.028)	(0.006)	(0.006)	(0.086)	(0.086)
Common last name	N	Y	N	Y	N	Y
Mean (sd) of DV	7.35 (1.88)	7.35 (1.88)	0.73 (0.44)	0.73 (0.44)	1.28 (5.43)	1.28 (5.43)
Observations	67,010	67,010	40,979	40,979	30,454	30,454
R-squared	0.56	0.56	0.09	0.09	0.04	0.04

Notes: OLS regressions. The unit of observation is a Scandinavian immigrant in the Unites States 1900 and 1910 population census. Commonness of first (and last) names is measured relative to the individual's birth cohort and gender in the earliest available Scandinavian census. Regressions include birth decade, decade of immigration, and Scandinavian census fixed effects along with controls for age in the US census, age squared, gender, number of first names, *log* county population, and *log* years since immigration to the US. Regressions in columns (3) and (4) are restricted to married immigrants and additionally controls for marriage after arrival to the US. Columns (5) and (6) include additional controls for the age, gender, and birth order of the oldest US-born child. Standard errors are double clustered at the Scandi. census × birth decade × birth country and household level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.16. Settlement of Scandinavian migrants in the United States: Robustness

Dependent var.	People of own nationality in county (*log*)		Spouse same nationality		Scandinavian name index of oldest US-born child	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Common names based on phonetic spelling						
Common first name	0.109***	0.109***	0.033***	0.033***	0.163**	0.164**
	(0.014)	(0.014)	(0.004)	(0.004)	(0.067)	(0.067)
Mean (sd) of DV	7.35 (1.88)	7.35 (1.88)	0.73 (0.44)	0.73 (0.44)	1.89 (4.74)	1.89 (4.74)
Observations	67,005	67,005	40,974	40,974	31,602	31,602
R-squared	0.55	0.55	0.08	0.09	0.04	0.04
Panel B: Common names within a ±1 birth year band						
Common first name	0.180***	0.180***	0.044***	0.043***	0.318***	0.318***
	(0.028)	(0.029)	(0.006)	(0.006)	(0.085)	(0.085)
Mean (sd) of DV	7.35 (1.88)	7.35 (1.88)	0.73 (0.44)	0.73 (0.44)	1.28 (5.43)	1.28 (5.43)
Observations	67,010	67,010	40,979	40,979	30,452	30,452
R-squared	0.56	0.56	0.09	0.09	0.04	0.04
Common last name	N	Y	N	Y	N	Y

Notes: The unit of observation is a Scandinavian immigrant in the Unites States 1900 and 1910 population census. Commonness of first (and last) names is measured relative to the individual's birth cohort and gender in the earliest available Scandinavian census. Panel A uses the phonetic spelling of first names and Panel B defines common first names within ±1 year band around the reported year of birth. Regressions include birth decade, decade of immigration, and Scandinavian census fixed effects along with controls for age in the US census, age squared, gender, number of first names, *log* county population, and *log* years since immigration to the US. Regressions in columns (3) and (4) are restricted to married immigrants and additionally controls for marriage after arrival to the US. Columns (5) and (6) include additional controls for the age, gender, and birth order of the oldest US-born child. Standard errors are double clustered at the Scandi. census × birth decade × birth country and household level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.17. Sum of cultural shocks 1860-1920

Country	District	Mean prevalence of common first names	Cultural shock of emigration	Cultural shock as a share of the mean prevalence	Cultural shock as a share of the within-country-SD in mean prevalence
Denmark	Aalborg	0.5279	0.0026	0.0049	0.0478
Denmark	Aarhus	0.4976	0.0033	0.0066	0.0608
Denmark	Bornholm	0.4538	0.0101	0.0224	0.1879
Denmark	Frederiksborg	0.5333	0.0051	0.0095	0.0936
Denmark	Hjørring	0.5061	0.0004	0.0009	0.0081
Denmark	Holbæk	0.6111	0.0027	0.0044	0.0501
Denmark	København	0.3995	0.0070	0.0176	0.1306
Denmark	Maribo	0.5484	0.0011	0.0020	0.0206
Denmark	Odense	0.6000	0.0025	0.0042	0.0469
Denmark	Præstø	0.6278	0.0012	0.0019	0.0226
Denmark	Randers	0.5282	0.0028	0.0052	0.0513
Denmark	Ribe	0.5267	0.0035	0.0066	0.0648
Denmark	Ringkøbing	0.5646	0.0031	0.0056	0.0581
Denmark	Roskilde	0.5758	0.0016	0.0027	0.0291
Denmark	Skanderborg	0.5151	0.0040	0.0077	0.0736
Denmark	Sorø	0.5896	0.0002	0.0003	0.0030
Denmark	Svendborg	0.5734	0.0050	0.0088	0.0933
Denmark	Thisted	0.5822	0.0026	0.0045	0.0488
Denmark	Vejle	0.5199	0.0029	0.0057	0.0546
Denmark	Viborg	0.5499	0.0025	0.0046	0.0468
Norway	Akershus	0.4063	0.0049	0.0121	0.1383
Norway	Aust-Agder	0.3275	0.0021	0.0064	0.0593
Norway	Bergen	0.2990	0.0093	0.0310	0.2606
Norway	Buskerud	0.3486	0.0077	0.0221	0.2167
Norway	Finnmark	0.3629	0.0325	0.0894	0.9118
Norway	Hedmark	0.3569	0.0032	0.0090	0.0907
Norway	Hordaland	0.4067	-0.0067	-0.0165	-0.1884
Norway	Møre and Romsdal	0.3803	-0.0022	-0.0059	-0.0629
Norway	Nord-Trøndelag	0.3556	0.0042	0.0117	0.1171
Norway	Nordland	0.3402	0.0042	0.0125	0.1191
Norway	Oppland	0.4009	-0.0017	-0.0042	-0.0469
Norway	Oslo	0.3289	0.0053	0.0161	0.1488
Norway	Rogaland	0.2961	0.0068	0.0231	0.1919
Norway	Sogn and fjordane	0.3727	0.0111	0.0299	0.3132
Norway	Sør-Trøndelag	0.3906	0.0026	0.0067	0.0740
Norway	Telemark	0.3693	0.0138	0.0375	0.3891
Norway	Troms	0.3489	0.0080	0.0229	0.2246
Norway	Vest-Agder	0.2953	-0.0002	-0.0008	-0.0065
Norway	Vestfold	0.3691	0.0105	0.0285	0.2956
Norway	Østfold	0.4021	0.0073	0.0182	0.2055
Sweden	Blekinge	0.4990	0.0033	0.0067	0.0742
Sweden	Gotland	0.5977	0.0254	0.0424	0.5630
Sweden	Gävleborg	0.6663	0.0267	0.0400	0.5916
Sweden	Göteborg och Bohus	0.5085	0.0107	0.0210	0.2369
Sweden	Halland	0.5739	0.0103	0.0179	0.2284
Sweden	Jämtland	0.6208	0.0183	0.0295	0.4069
Sweden	Jönköping	0.6103	0.0267	0.0437	0.5921
Sweden	Kalmar	0.6018	0.0307	0.0510	0.6818
Sweden	Kopparberg	0.6241	0.0189	0.0303	0.4191
Sweden	Kristianstad	0.5530	0.0029	0.0053	0.0645
Sweden	Kronoberg	0.6013	0.0169	0.0281	0.3750
Sweden	Malmöhus	0.5646	0.0091	0.0160	0.2011
Sweden	Norrbotten	0.5907	0.0147	0.0249	0.3261
Sweden	Skaraborg	0.6475	0.0231	0.0357	0.5134
Sweden	Stockholm	0.5557	0.0148	0.0266	0.3283
Sweden	Södermanland	0.6586	0.0094	0.0143	0.2096
Sweden	Uppsala	0.6403	0.0083	0.0130	0.1843
Sweden	Värmland	0.6158	0.0335	0.0543	0.7425
Sweden	Västerbotten	0.5891	0.0072	0.0121	0.1588
Sweden	Västernorrland	0.6022	0.0190	0.0315	0.4210
Sweden	Västmanland	0.6777	0.0189	0.0278	0.4184
Sweden	Älvsborg	0.5797	0.0183	0.0316	0.4067
Sweden	Örebro	0.6474	0.0220	0.0340	0.4884
Sweden	Östergötland	0.6075	0.0221	0.0363	0.4899

Notes: Data for Figure 3.

Table D.18. Persistence in the short run, 1860-1910: Alternative shock measures

Dependent variable	Share of newborns with common first names				
	(1)	(2)	(3)	(4)	(5)
Panel A	OLS estimation				
Common first names,	0.895***	0.889***	0.889***	0.890***	0.898***
sum of lags	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Cultural shock of emi.,	1.213**	1.287**	1.309**	1.273**	2.792***
first lag	(0.565)	(0.529)	(0.532)	(0.485)	(0.770)
R-squared	0.98	0.98	0.98	0.99	0.99
Panel B	Fixed effects estimation				
Common first names,	0.626***	0.626***	0.627***	0.626***	0.623***
sum of lags	(0.045)	(0.045)	(0.045)	(0.045)	(0.041)
Cultural shock of emi.,	1.281*	1.000*	1.023*	1.014*	3.356***
first lag	(0.653)	(0.578)	(0.585)	(0.549)	(0.917)
R-squared	0.96	0.96	0.96	0.96	0.97
Panel C	Difference GMM estimation				
Common first names,	0.711***	0.707***	0.707***	0.706***	0.705***
sum of lags	(0.038)	(0.039)	(0.039)	(0.039)	(0.037)
Cultural shock of emi.,	1.397**	1.072*	1.093*	1.059**	3.416***
first lag	(0.646)	(0.563)	(0.567)	(0.518)	(0.768)
AR2 test *p*-value	0.729	0.793	0.802	0.794	0.913
Cultural shock measure	Phonetic spelling of names	Common names within ±1 birth year band	Excl. emigrants w. name not in census	Cohorts aged 0-50	Unweighted emigrants
Mean (sd) of DV	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)
Observations	320	320	320	320	320

Notes: Dynamic panel results. The dependent variable is measured across 64 subnational districts in the decades 1860 to 1910. *Cultural shock of emigration* measures the increase in the share of common first names in the district population due to selective emigration. This is based on different calculations described in Appendix B.2. All regressions include decade fixed effects and two lags of the dependent variable. Controls include the log number, gender ratio, and mean number of first names of people born in the decade and district, along with the log number of emigrants (first lag). The panels contain no unit roots (Levin-Lin-Chu *t*-stat -9.420, *p*-value 0.000). Robust standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.19. Persistence of cultural shocks in the medium run

	(1)	(2)	(3)	(4)	(5)
Panel A	Dep. var.: Common first names 1901/1910				
Common first names 1845/65/80	0.459***	0.468***	0.678***	0.474***	0.528***
	(0.060)	(0.060)	(0.073)	(0.072)	(0.068)
Cultural shock of the AoMM	1.840***	1.773***	0.900*	1.968***	1.291*
	(0.565)	(0.596)	(0.494)	(0.576)	(0.650)
Mean (sd) of DV	.371 (.087)	.371 (.087)	.371 (.087)	.371 (.087)	.371 (.087)
R-squared	0.92	0.92	0.94	0.93	0.93
Panel B	Dep. var.: Extended families 1901/1910				
Extended families 1845/65/80	0.479***	0.440***	0.470***	0.422***	0.514***
	(0.070)	(0.053)	(0.058)	(0.046)	(0.073)
Cultural shock of the AoMM	3.259***	2.870***	2.466***	1.673*	1.536**
	(0.843)	(0.652)	(0.586)	(0.999)	(0.749)
Mean (sd) of DV	.498 (.079)	.498 (.079)	.498 (.079)	.498 (.079)	.498 (.079)
R-squared	0.50	0.67	0.69	0.79	0.72
Panel C	Dep. var.: Intra-marriage 1901/1910				
Intra-marriage 1845/65/80	1.489***	1.304***	1.260***	1.324***	1.161***
	(0.279)	(0.349)	(0.343)	(0.337)	(0.328)
Cultural shock of the AoMM	1.515**	1.325*	0.869	0.182	0.833
	(0.726)	(0.733)	(0.765)	(0.742)	(0.757)
Mean (sd) of DV	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)
R-squared	0.82	0.83	0.83	0.86	0.85
Observations	64	64	64	64	64
Baseline controls	Y	Y	Y	Y	Y
Log emigration	N	Y	Y	Y	Y
Common last names	N	N	Y	N	N
Geography	N	N	N	Y	N
Country fixed effects	N	N	N	N	Y

Notes: The unit of observation is a subnational district. The dependent variable is the prevalence of collectivist norms at the end of the Age of Mass Migration, observed the most recent census from each country (1901 in DK and 1910 in NO and SE). In Panel A this equals the share of children ($\leq$ 10 years old) with common first names, in Panel B the share of elderly people ($\geq$ 65 years old) who live with relatives from younger generations (*extended families*), and in Panel C the share of married couples between the ages 20-40 who are born in the same district (*intra-marriage*). The *1845/65/80* measures of the same norms are taken from the earliest census from each country (1845 in DK, 1865 in NO, and 1880 in SE). *Cultural shock of the AoMM* measures the increase in the share of common first names in the district population directly due to selective emigration in the Age of Mass Migration. All regressions control for log population in the earliest census along with the children's gender ratio and mean number of first names in Panel A, the share of elderly in the population in Panel B, and the share of people age 20 to 40 who live outside their district of birth in Panel C. Geographical controls include log area, log distance to the capital, the nearest emigration port, and the coast. Robust standard errors in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.20. Persistence in the medium run: Alternative shock measures

	(1)	(2)	(3)	(4)	(5)
Panel A	Dep. var.: Common first names in 1901/1910				
Common first names 1845/65/80	0.469***	0.458***	0.458***	0.457***	0.479***
	(0.068)	(0.059)	(0.060)	(0.057)	(0.062)
Cultural shock of the AoMM	1.363*	1.923***	1.849***	1.726***	1.890**
	(0.723)	(0.621)	(0.636)	(0.487)	(0.787)
Mean (sd) of DV	.371 (.087)	.371 (.087)	.371 (.087)	.371 (.087)	.371 (.087)
R-squared	0.91	0.92	0.92	0.92	0.91
Panel B	Dep. var.: Extended families in 1901/1910				
Extended families 1845/65/80	0.436***	0.441***	0.442***	0.440***	0.428***
	(0.054)	(0.053)	(0.054)	(0.053)	(0.054)
Cultural shock of the AoMM	3.055***	2.846***	2.938***	2.369***	3.289***
	(0.762)	(0.696)	(0.694)	(0.619)	(0.969)
Mean of DV	.498 (.079)	.498 (.079)	.498 (.079)	.498 (.079)	.498 (.079)
R-squared	0.66	0.67	0.67	0.67	0.66
Panel C	Dep. var.: Intra-marriage in 1901/1910				
Intra-marriage 1845/65/80	1.315***	1.284***	1.279***	1.289***	1.286***
	(0.349)	(0.348)	(0.345)	(0.350)	(0.353)
Cultural shock of the AoMM	1.652**	1.220	1.326*	0.961	1.024
	(0.809)	(0.766)	(0.771)	(0.669)	(1.026)
Mean of DV	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)	.698 (.119)
R-squared	0.83	0.83	0.83	0.83	0.82
Cultural shock measure	Phonetic spelling of names	Common names within ±1 birth year band	Excl. emigrants w. name not in census	Cohorts aged 0-50	Unweighted emigrants
Observations	64	64	64	64	64

Notes: The unit of observation is a subnational district. The dependent variable is the prevalence of collectivist norms at the end of the Age of Mass Migration, observed the most recent census from each country (1901 in DK and 1910 in NO and SE). In Panel A this equals the share of children ($\leq$ 10 years old) with common first names, in Panel B the share of elderly people ($\geq$ 65 years old) who live with relatives from younger generations (*extended families*), and in Panel C the share of married couples between the ages 20-40 who are born in the same district (*intra-marriage*). The *1845/65/80* measures of the same norms are taken from the earliest census from each country (1845 in DK, 1865 in NO, and 1880 in SE). *Cultural shock of the AoMM* measures the increase in the share of common first names in the district population directly due to selective emigration in the Age of Mass Migration. This is based on different calculations described in Appendix B.2. Controls include log initial population and log emigration, along with the children's gender ratio and mean number of first names in Panel A, the share of elderly in the population in Panel B, and the share of people age 20 to 40 who live outside their district of birth in Panel C. Robust standard errors in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.21. Predicting contemporary collectivist attitudes (WVS/EVS): Additional control variables

Dependent var.	Important child characteristics:						Spatial identity:			Trust in:		
	Obedience vs. independence			+ unselfishness vs. determination			Local vs. national or global			Family vs. others		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Common first names 1845/65/80	-0.031	0.111	0.017	0.423**	0.396**	0.573**	0.033	0.281***	-0.068	0.004	0.154	0.149
	(0.210)	(0.138)	(0.225)	(0.205)	(0.155)	(0.216)	(0.166)	(0.103)	(0.079)	(0.266)	(0.180)	(0.218)
Cultural shock of the AoMM	7.010***	6.114**	6.943**	6.019***	5.402***	7.262***	3.403***	2.921*	3.593***	7.550**	5.663*	7.081**
	(2.569)	(2.464)	(2.855)	(1.806)	(1.774)	(2.015)	(1.242)	(1.585)	(1.197)	(2.880)	(3.099)	(2.925)
Baseline controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Log emigration (district)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Common last names (district)	Y	N	N	Y	N	N	Y	N	N	Y	N	N
Marriage & education (respondent)	N	Y	N	N	Y	N	N	Y	N	N	Y	N
Geography (district)	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Mean (sd) of DV	standardized			standardized			0.185 (0.388)			standardized		
No. of districts	50	50	50	50	50	50	50	50	50	50	50	50
Observations	13,003	12,756	13,003	13,003	12,756	13,003	10,959	10,771	10,959	7,847	7,687	7,847
R-squared	0.04	0.07	0.04	0.05	0.07	0.05	0.07	0.08	0.08	0.10	0.11	0.10

Notes: The unit of observation is an individual respondent in the pooled WVS and EVS 1996-2018. Explanatory variables are measured in the respondents' district of residence (NUTS 3 level): *Common first names* equals the share of children with common first names in 1845 in DK, 1865 in NO, and 1880 in SE. *Cultural shock* is the sum of cultural shocks of emigration from 1845/65/80 until the end of the Age of Mass Migration (AoMM) in 1920. Baseline controls include the *log* population, child gender ratio, and the mean number of first names among children in 1845/65/80. Common last names are measured among children in 1845/65/80. Geographical controls include log area, log distance to the capital, the nearest emigration port, and the coast. All regressions include country and WVS/EVS wave fixed effects along with individual respondent controls for age, age squared, and gender. Robust standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.22. Predicting contemporary attitudes (WVS/EVS): Alt. shock measures

	(1)	(2)	(3)	(4)	(5)
Panel A	Dep. var.: Obedience vs. independence (child characteristics)				
Common first names	0.330**	0.314*	0.320*	0.303*	0.454**
1845/65/80	(0.163)	(0.161)	(0.163)	(0.163)	(0.174)
Cultural shock of	8.116**	6.482**	6.546**	5.230**	4.273
the AoMM	(3.067)	(2.878)	(3.007)	(2.474)	(3.247)
Observations	13,003	13,003	13,003	13,003	13,003
R-squared	0.04	0.04	0.04	0.04	0.04
Panel B	Dep. var.: + unselfishness vs. determination (child characteristics)				
Common first names	0.599***	0.582***	0.586***	0.563***	0.728***
1845/65/80	(0.170)	(0.165)	(0.166)	(0.168)	(0.158)
Cultural shock of	6.607***	5.765***	5.923***	5.166***	5.104**
the AoMM	(2.159)	(1.947)	(2.050)	(1.653)	(2.361)
Observations	13,003	13,003	13,003	13,003	13,003
R-squared	0.05	0.05	0.05	0.05	0.05
Panel C	Dep. var.: Local spatial identity				
Common first names	0.335***	0.331***	0.333***	0.326***	0.405***
1845/65/80	(0.097)	(0.099)	(0.100)	(0.100)	(0.089)
Cultural shock of	3.141	2.973*	2.939*	2.315*	2.774
the AoMM	(2.022)	(1.566)	(1.633)	(1.358)	(1.828)
Mean (sd) of DV	.185 (.388)	.185 (.388)	.185 (.388)	.185 (.388)	.185 (.388)
Observations	10,959	10,959	10,959	10,959	10,959
R-squared	0.07	0.07	0.07	0.07	0.07
Panel D	Dep. var.: Trust in family vs. others				
Common first names	0.323	0.331*	0.340*	0.319	0.518***
1845/65/80	(0.195)	(0.191)	(0.193)	(0.194)	(0.189)
Cultural shock of	11.769***	6.981**	6.714**	5.632**	6.540*
the AoMM	(3.934)	(3.110)	(3.305)	(2.676)	(3.590)
Observations	7,847	7,847	7,847	7,847	7,847
R-squared	0.10	0.10	0.10	0.10	0.10
Cultural shock measure	Phonetic spelling of names	Common names within ±1 birth year band	Excl. emigrants w. name not in census	Cohorts aged 0-50	Unweighted emigrants

Notes: The unit of observation is an individual respondent in the pooled WVS and EVS 1996-2018. See the main text for a description of dependent variables. Explanatory variables are measured in the respondents' district of residence (NUTS 3 level): *Common first names* equals the share of children with common first names in 1845 in DK, 1865 in NO, and 1880 in SE. *Cultural shock* is the sum of cultural shocks of emigration from 1845/65/80 until the end of the Age of Mass Migration (AoMM) in 1920. This is based on different calculations described in Appendix B.2. Controls include the *log* population, child gender ratio, and the mean number of first names among children in 1845/65/80, along with the log of total emigration. All regressions include country and WVS/EVS wave fixed effects along with individual respondent controls for age, age squared, and gender. Robust standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.23. Predicting contemporary naming patterns

Dependent var.	Common first names 2010s				
	(1)	(2)	(3)	(4)	(5)
Common first names 1845/65/80	-0.023	-0.007	0.137*	-0.069	0.055
	(0.080)	(0.080)	(0.073)	(0.044)	(0.035)
Cultural shock of the AoMM	1.798***	1.661***	0.909*	0.873*	0.619*
	(0.440)	(0.498)	(0.472)	(0.466)	(0.344)
Baseline controls	Y	Y	Y	Y	Y
Log emigration	N	Y	Y	Y	Y
Common last names	N	N	Y	N	N
Geography	N	N	N	Y	N
Country fixed effects	N	N	N	N	Y
Mean of DV	0.136	0.136	0.136	0.136	0.136
SD of DV	0.024	0.024	0.024	0.024	0.024
Observations	50	50	50	50	50
R-squared	0.41	0.41	0.63	0.61	0.78

Notes: The unit of observation is a subnational district (NUTS 3 level). The dependent variable is the share of newborns born in the 2010s with common first names. *Common first names 1845/65/80* is the share of children with common first names in 1845 in DK, 1865 in NO, and 1880 in SE. *Cultural shock* is the sum of cultural shocks of emigration from 1845/65/80 until the end of the Age of Mass Migration (AoMM) in 1920. Baseline controls include the *log* population, child gender ratio, and the mean number of first names among children in 1845/65/80. Common last names are measured among children in 1845/65/80. Geographical controls include log area, log distance to the capital, the nearest emigration port, and the coast. Robust standard errors in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.