Those Who Stayed: Selection and Cultural Change during the Age of Mass Migration*

Anne Sofie Beck Knudsen†

Job Market Paper

November, 2019

Abstract

This paper studies the cultural causes and consequences of mass emigration from Scandinavia in the 19th century. I test the hypothesis that people with individualistic traits were more likely to emigrate, because they faced lower costs of leaving established social networks behind. Data from population censuses and passenger lists confirm this hypothesis. Children who grew up in households with nonconformist naming practices, nuclear family structures, and weak ties to parents' birthplaces were on average more likely to emigrate later in life. Selection was weaker under circumstances that reduced the social costs of emigration. This was the case with larger migration networks abroad, and in situations where people emigrated collectively. Based on these findings, I expect emigration to generate cultural change towards reduced individualism in migrant-sending locations, through a combination of initial compositional effects and intergenerational cultural transmission. This is confirmed in a cross-district setting with measures of actual cultural change over the medium and long run.

Keywords: Migration, selection, culture, individualism, Age of Mass Migration
JEL classification: Z10, F22, O15, R23, N33

*I am grateful to Carl-Johan Dalgaard for his encouragement and advice on this project. I also thank Jeanet Sinding Bentzen, Sascha Becker, Carl-Johan Dalgaard, Vasiliki Fouka, Oded Galor, Daniel Gerszon Mahler, Nathan Nunn, Lena Sperling, Uwe Sunde, and seminar participants at Brown, Harvard, Lund, Southern Denmark, Stanford, UC Davis, the NBER Summer Institute, and ASREC conferences for helpful comments and suggestions.

†Department of Economics, Harvard University, email: aknudsen@fas.harvard.edu

1 Introduction

Voluntary migration is a selective process where some people choose to leave and some choose to stay in their location of origin. A question is then, who migrates? Research has documented that migrants self-select for labor market success on the basis of more or less productive skills (Borjas 1987, Abramitzky and Boustan 2016). But selection likely takes place on other dimensions as well. There is psychological evidence that the social and existential costs of migration vary systematically across individuals with individualistic or collectivist cultural traits.[1] A hypothesis proposes that people with individualistic traits are thus more likely to migrate (Kitayama et al. 2006). Being less reliant on others for support and identity, they face a lower costs of leaving established social networks behind. If true, and if culture is transmitted across generations, then migration may generate cultural change towards collectivism in the migrant-sending locations, simply through a change in the composition of cultural types. The societal implications of such a relationship can be wide-reaching as differences in individualism and collectivism has been linked to economic development through their impact on innovation, cooperation, and public goods provision.[2]

In this paper, I study the relationship between individualistic cultural traits and migration. A priori it is not clear if the hypothesis of individualistic migration can be applied to all migration situations. For instance, one would expect the social costs of migration to decrease when migrant network abroad grow. Social costs may even become negative if a large share of the migrants network from home migrate as well. I therefore not only test if selection on individualistic traits take place on the average. I also test if selection weakened by circumstances that would reduce the social costs of migration. To examine these questions empirically, I take a historical perspective and study the Scandinavian experience in one of the largest migration events in modern history, the Age of Mass Migration in the years 1850 to 1920. Here, I find that Scandinavians who grew up in individualistic households were on average more likely to emigrate. As expected selection weakened as emigration networks abroad grew and in situations where emigration happened collectively. Studying the effects of selective emigration across migrant-sending locations, I find that emigration was associated with a reduction in the prevalence of individualistic traits, which lasted over time. Cross-district empirical analysis reveal that the cultural change that took place during the Age of Mass Migration was sufficiently profound to leave a long-run impact on contemporary Scandinavian culture.

Several circumstances make the Age of Mass Migration an ideal case for the empirical

[1] Studies on immigrants' well-being find that feelings of social and existential loss are more prevalent among migrants from collectivist, compared to individualistic, societies (Bhugra and Becker 2006).

[2] See Morris et al. (1993), Greif (1994), Wagner III (1995), Chen et al., (1998), van Everdingen and Waarts (2003), Bozeman (2007), Gorodnichenko and Roland (2011a), Taylor and Wilson (2012), and Buggle (2017).

objective of this paper. During the period, millions of people left Europe to settle in New World countries such as the United States. Emigration was fascilitated by the drop in transport costs associated with the shift from sail to steam shipping technology and the end of the American Civil War in 1865. Sweden, Norway, and Denmark experienced some of the highest emigration rates in Europe during this period, involving the departure of approximately 25% of their populations. Global regulatory policies on migration were particularly loose at this point in time, which enables the identification of self-selective processes under limited governmental influence. In addition, the historical context allows me to study long-run cultural implications of migration in sending locations.

An empirical challenge in research on cultural traits is the lack of historical data. Drawing on research in the field of social psychology, I construct indicators of collectivism versus individualism based on the commonness of first names, which correlate with a desire to fit in rather than stand out (Zweigenhaft, 1981; Twenge et al., 2010; Emery, 2013). Properties of a first name choice reflect the preferences of the name-givers, and the commonness of a first name thus measures the collectivist traits inherited from home. The predetermined nature of first names provides a useful source of identification. I validate the measure using other, both contemporary and historical, indicators of individualism.

With proxies of collectivism and individualism in hand, I proceed to empirically examine the relationship between migration and individualistic cultural traits. First, I test the hypothesis that people with individualistic traits are more likely to migrate, and that people with collectivist traits hence more likely to stay. For this I construct a novel database of migrants and non-migrants who lived in Scandinavia just before or during the Age of Mass Migration. The Scandinavian countries were particular meticulous in registering not just their home populations but everyone that left to settle elsewhere. I am thereby able to combine historical population census records with detailed passenger lists from ships that carried migrants abroad. The result is a database of 1,253,317 Scandinavian first-time emigrants that cover 70% of the total emigration flows of the period. For a subset of these emigrants, I identify their childhood households in the population census records. Analysis of the data confirms that individuals who grew up in collectivist households were less likely to emigrate. This remains true with controls for other circumstances that may have shaped the cultural, social, and economic prospects of young individuals such as birth order and parent's socioeconomic status. The results also hold with alternative proxies of collectivist values based on marriage and extended family patterns.

Not all emigrants were equally selected. Some had more common and some more uncommon first names. Looking at the timing of emigration, I find that time-varying district level circumstances mattered for the cultural selection into emigration. In decades with larger emigrant outflows, with a higher cumulative stock of emigrants abroad, and with more severe weather shocks, selection was weaker. The underlying intuition behind these results is that these circumstances reduced the relative social costs of emigration,

prompting more collectivist individuals to join the emigration flows.

The act of migration can be characterized in a number of directions. Having seen that time-varying circumstances at home impact the strength by which Scandinavian emigrants selected on individualistic cultural traits, a natural question then is if cultural heterogeneity across emigrants explains other aspects of the migration decision. Analyses of the behavior of emigrants reveal that emigrants were more likely to emigrate with family members or fellow townspeople. They were also more likely to seek the company of fellow Scandinavians and preserve customs from home, once arriving at their destination. These results support the idea that people with common first names valued established social connections, and that this impacted their emigration behavior.

For use in analyses of cross-district ultural change, I quantify the cultural shock of selective emigration. This reflects the relative increase in the prevalence of collectivism in the migrant-sending population that occurs due to emigration. Approximating this is feasible because I have information on rates of emigration and the gap between emigrant and population collectivism. With this, I assess the cultural impact of the Age of Mass Migration on medium and long run cultural differences in Scandinavia.

I find that collectivistic traits before the Age of Mass Migration *and* the increase in collectivistic traits associated with selective emigration significantly predict subsequent measures of collectivism and individualism. First, this is is true for the prevalence of common first names towards the end of the Age of Mass Migration around 1910. Second, it is true for the prevalence of common first names in 2015. Finally, it is true when using survey responses from the World Values Survey and European Values Study from the period 1981-2012. Importantly, the results hold when controlling for the total rate of emigration over the period to ensure that it is in fact via its initial compositional impact, and not its magnitude alone, that emigration matters for cultural change. Although this evidence is more descriptive, it suggests that subnational cultural differences persisted over a period of up to 170 years, and that the cumulative sum of emigration shocks was associated with a substantial decline in contemporary individualism.

Taken together, the empirical results of this paper show that measures of collectivism exhibited significant but imperfect persistence across localities from before the Age of Mass Migration and until today. Cultural change took place as people with individualistic cultural traits selected into emigration and left Scandinavia. I thereby draw attention to determinants and effects of migration that have not been studied before but are of broad relevance.

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 motivates and describes the main data sources and empirical measures. Section 4 tests for selective emigration using individual level data, and section 5 studies the implications of selection on district-level cultural evolution. Section 6 concludes.

Related Literature This paper provides the most comprehensive test of the *voluntary settlement hypothesis* on selective migration by Kitayama et al. (2006) so far produced. The hypothesis was formed as a version of the *frontier hypothesis* that American individualism was the result of a long history of living on an expanding frontier (Turner, 1920). Instead, emphasis was put on the voluntary movement of particularly autonomous and independent individuals into frontier regions, and how this self-selection shaped subsequent cultural developments (Kitayama et al., 2006; Varnum and Kitayama, 2010). To test the hypothesis, I narrow the focus on the individualistic selection of migrants and examine the phenomenon in the original home locations of the migrants. I also investigate its consequences for cultural change in migrant-sending localities, which represents a natural yet unexplored extension of the hypothesis.

The results of this paper contribute more generally to the understanding of determinants and consequences of migration. Studies of migrant selection have traditionally focused on the impact of economic circumstances (see Abramitzky et al., 2017, for review). Of particular relevance is the work by Abramitzky et al. (2013) that documents how historical Norwegian emigrants were negatively selected in terms of economic prospects at childhood. An exception is Jaeger et al. (2010) who find evidence of a link between risk attitudes and migration, based on ex post characteristics of migrants and stayers in the contemporary German population. The literature on effects of emigration on migrant-sending communities is also related to this study. This includes the literature on modern globalization and brain drain (Docquier and Rapoport, 2012) and recent studies that link historical Swedish mass emigration to the rise of labor unions and innovative activity across Swedish localities (Anderson et al., 2017; Karadja and Prawitz, 2019).

Another relevant strand of literature is the field on long run development, which is concerned with factors that have ultimately shaped the world as it is today. Among other factors, this literature focuses on the role played by culture, and it is identifying the historical origins of contemporary cultural differences and their impact on economic and institutional developments. This includes a growing number of studies on individualism (Greif, 1994; Gorodnichenko and Roland, 2011a, 2011b, 2017; Olsson and Paik, 2016; Buggle, 2017; Bazzi et al., 2018). There are less studies on the processes of cultural change (with examples being Voigtländer and Voth, 2012; Guiliano and Nunn, 2018; Fouka, 2019; and Bentzen, 2019).[3] This paper contributes to this literature by providing insights on a specific channel of cultural change, migration, and its reciprocal relationship with the cultural trait itself in the short as well as the long run.

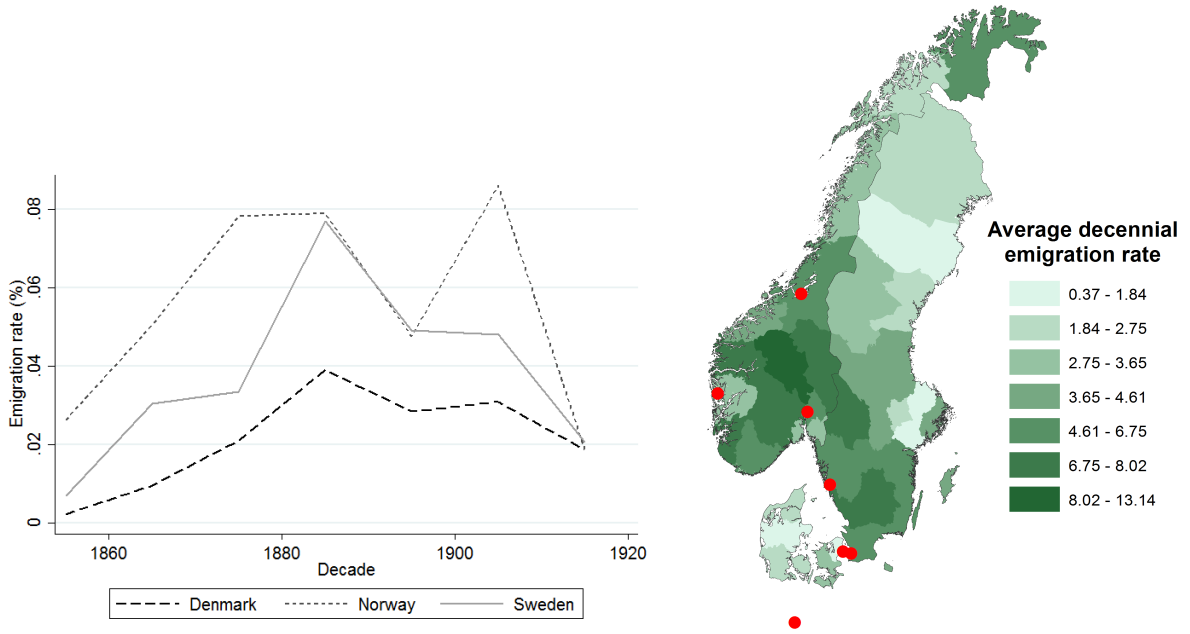
[3]Guiliano & Nunn (2019) show that locations subject to climatic instability experience more rapidly changing cultures. Fouka (2019) documents that the implementation of forced assimilation policies in the US impacted the cultural integration of German immigrants negatively, and Bentzen (2019) finds that religiosity increases in locations immediately after being hit by earthquakes. When studying the long term persistence of anti-Semitic beliefs, Voigtländer & Voth (2012) show that growing city size and trade weakens persistence.

2 Background

2.1 Historical Context

Between 1850 and 1920 more than 40 million Europeans emigrated to the New World, with the majority settling in the United States (Taylor and Williamson, 1997). Scandinavia experienced some of the highest emigration rates in Europe during the time period. Migration was facilitated by a near absence of regulatory migration policies and the shift from wind to steam driven ship transportation in the late 1860's. According to the numbers in Figure 1, emigration peaked in the 1880's where around 8% of the Swedish and Norwegian populations left in a single decade. Although Danish emigration flows were weaker, nearly 15% of the population left altogether. Total emigration amounted to around 38% and 26% in Norway and Sweden respectively - numbers only matched by Ireland and Italy (Hatton and Williamson, 1992). The intensity of emigration varied considerably across Scandinavian districts as can be seen from the map in Figure 1, where major emigration ports are also shown.

Figure 1. Scandinavian emigration 1850-1920



Notes: Decennial rates of emigration over time (left) and across districts (right). See section 3.3 for a description of data sources. Districts represent the current subnational divisions and refer to 21 Swedish counties (län), 19 Norwegian counties (pre-2018 fylke), and 10 Danish provinces (landsdele). Major emigration ports for the Scandinavian emigrants (the red dots) include Gothenburg, Malmö, Oslo, Trondheim, Bergen, Copenhagen, and Hamburg.

In related empirical research, the causes of migration during this time period have especially been linked to economic circumstances. Years of failed harvests, structural transformation, and economic inequality generated a push of particularly low skilled and agricultural workers who benefitted from better job opportunities and access to cheap

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). Stories on the '*American Dream'* and the view of the United States as the *'Land of Opportunities'* were core to the migration discourse. Private letters, diaries, and newspaper articles of the time reveal that ideas of personal freedom and social equality embodied in the American society were of great value to the migrants. In the United States, people were free to pursue own goals. The countries of the New World thereby represented a contrast to Scandinavia that, besides offering poor economic prospects, were characterized by social rigidity, religious intolerance, and limited suffrage.[4] The historical evidence suggests that cultural motivations were part of the migration decision in addition to economic incentives.

2.2 Conceptual Considerations

The act of migration involves leaving familiar and established social networks behind. Depending on the value placed on these networks, and circumstances at destination, the migration experience is potentially associated with great psychological distress (Eisenbruch, 1991; Bhugra and Becker, 2005). This observation has motivated research to study non-economic drivers and consequences of migration. In social psychology, one of the more prominent theories, the *voluntary settlement hypothesis*, proposes that people of a stronger individualistic cultural beliefs are more likely to migrate and settle somewhere new than their collectivist counterparts (Kitayama et al., 2006).

An individualistic culture is commonly defined as centering on the belief that human beings constitute autonomous units in loosely-knitted social networks. Here, individual identity is derived from inner attributes, abilities, and personality traits. In contrast, a collectivist culture 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 includes the emphasis of distinction over conformity or the pursuit of individual versus collective goals in individualistic and collectivist cultures, respectively (Heine and Ruby, 2010).

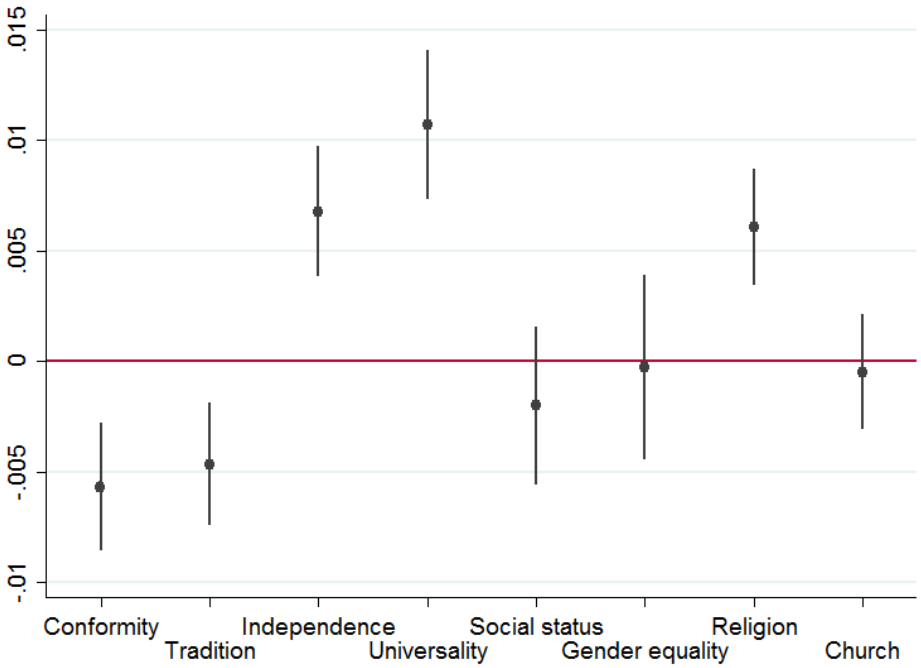
Based on these characteristics, it seems likely that this cultural trait plays an important role in the migration decision. First, people of an individualistic mindset suffer a lower cost of abandoning existing social networks - both in terms of loss of identity and support system. Second, they may experience a personal satisfaction from doing something out of the ordinary that requires and shows independent effort. While migration is certainly driven by other factors than individualism and collectivism, one would expect

[4]People of minority religious denominations were for instance among the first to emigrate, including Danish and Norwegian Mormons and members of the Swedish free church movement. Some were fleeing actual religious persecution.

individualists to respond more readily in a setting of various push and pull effects.

Figure 2 presents some motivating evidence in support of this hypothesis. The European Social Study is the only large-scale database that contain information on both migration status and cultural values. Comparing European migrants with non-migrants reveal significant cultural differences in values associated with collectivism and individualism. Migrants are more likely to value independence and universal moral values than tradition and conformity. There are no significant differences in the value of social status or gender equality. Although emigrants tend to emphasize religious values as well, there are no significant differences in their propensity to go to church. A major caveat of this data is that cultural values are recorded post-migration, and they may be the outcome of a transformative emigration experience. The empirical analysis of this paper uses pre-determined cultural proxies to overcome this issue.

Figure 2. Cultural values and the propensity to migrate within Europe, 2002-2006



Notes: Standardized coefficients from regressing international migrant status on individual values and beliefs in the European Social Survey, 2002-2016 (N=334,541). Migrant status is defined as living outside of one's country of birth. Regressions control for age, age squared, marriage, education, birth year, and fixed effects for survey year fixed effect, country of origin, and country of residence. Bars represent 95% confidence intervals. Included values are values of obedience and proper behavior (conformity), tradition and modesty (tradition), freedom to pursue own goals (independence), universal vs. communal morality (universality), earning respect and admiration from others (social status), women and men having the same right to work (gender equality), religiosity (religion), and the frequency of church going (church going).

If people migrate based, in part, on individualistic cultural values, migration will have implications on the overall evolution of cultures. Emigration must be associated with an immediate reduction in the prevalence of individualistic cultural traits in the migrant-sending population. This is a simple composition effect that I refer to as the *cultural*

shock of emigration. The magnitude of the shock depends on the rate and selection 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, so-called role models, of society (Cavalli-Sforza and Feldman, 1981; Bisin and Verdier, 1998, 2000, 2001; Dohmen et al., 2012; Chowdhury et al. 2018). In a related study I document the presence of these channels in the cultural transmission of individualistic versus collectivist cultural values in the same historical setting as the present study (Knudsen, 2018).

With channels of intergenerational cultural transmission in place, emigration is expected to push the cultures of migrant-sending locations in a relatively more collectivist direction. The intensity with which this happens need not be proportional to the initial shock. The shock may alter both the supply of cultural role models and the benefits from being one or the other cultural type. An example is if parents of distinct cultural values become a minority.[5] Such mechanisms may strengthen or weaken the initial impact of the shock.

Another expectation is that the intensity of cultural selection will weaken over time. When migrant networks abroad are created and a common acceptance and experience with emigration spreads at home, the social cost of migration diminishes and more collectivists are expected to join the migration flows. The same will be true when non-cultural push and pull factors are strong, which would include adverse weather shocks.

The above considerations may be summarized in three testable hypotheses:

Hypothesis 1 (cultural selection). *People that emigrated in the Age of Mass Migration were more individualistic than the people that stayed behind.*

Hypothesis 2 (moderators of selection). *The cultural selection of emigrants weakened when other push and pull factors, such as migrant networks abroad and weather shocks at home, strengthened.*

Hypothesis 3 (cultural change). *The reduction in home population individualism directly associated with selective emigration persisted over time through channels of intergenerational cultural transmission.*

[5]Parents of distinct cultural values may alter their socialization efforts when the cultural environment changes and they become a cultural minority. On the one hand minority individualists may be more determined in ensuring the survival of their (individualistic) traits. On the other hand, one could imagine a social cost of being individualist in an increasingly collectivist environment, whereby more individualists may become collectivists. Likewise, finding a spouse that shares one's individualistic traits can be more difficult.

3 Data Sources and Measurement

3.1 Population Census Records

One of two main data sources for this paper is a set of historical censuses on the Scandinavian population over the period 1845-1910. I benefit from the ongoing work by national statistical offices and academic institutions with digitizing these archives.[6] The digital copies of the Swedish and Norwegian censues are of particular high quality with numerical codes on occupations, locations, and interrelationships between household members. I assign these codes to the Danish census data based on descriptions given by the enumerators at the time of data collection. Of the digitized population census records, I use those on entire populations that contain information on residence and birthplace district. The resulting set of censuses cover the years 1845, 1880, and 1901 for Denmark, 1880, 1890 and 1900 for Sweden, and 1865, 1900, and 1910 for Norway. Besides documenting demographic and economic circumstances of all individuals and their respective households, the censuses include information on full names from which cultural indicators can be constructed (see section 3.3).

The censuses report the location of residence and birth of all individuals. I use this information to link individuals across data sources and to construct geographically narrow birth cohorts. My preferred level of subnational aggregation is at the historical district level. These refer to 20 *fylke* in Norway, 24 *län* in Sweden, and 23 *amter* in Denmark. Although the borders of these districts did not change during the historical time period considered in the paper, they have changed since then. The analysis in section 5, where I include contemporary survey data, is therefore conducted at the level of modern districts, which are fewer than the historical districts.[7]

3.2 Emigration Data

I construct two datasets on the subnational emigration patterns in Scandinavia during the Age of Mass Migration: One with information at the individual migrant level and one at the subnational district level.

[6] The Swedish and Norwegian data is made available by the North Atlantic Population Project at the Minnesota Population Center (2017), while the Danish data is obtained through the Danish National Archives.

[7] The modern NUTS-2 districts are illustrated in Figure 1. The biggest change took place in Denmark where the historical districts have been collapsed to 10 bigger regions. In Norway, one historical district (Bergen) is now included in the larger Hordaland district, and in Sweden 5 historical districts were merged into two (Malmöhus and Kristianstad län became Skåne; Älvsborg, Skaraborg, and Gothenburg and Bohus became Västra Götaland).

3.2.1 A Database of Scandinavian Emigrants

I use detailed passenger lists to create a database of a majority of the emigrants that left Scandinavia in this time period. In 1868 and 1869 national laws were passed in each Scandinavian country, requiring travel agencies to record personal details of all passengers who left on ships bound for international destinations. These laws were inspired by a similar system in Hamburg and implemented to ensure the proper conduct of passengers' money and onboard safety. The quality of this information can be trusted to be high as it was reported to the national police, whose control and approval was required for the agencies to continue their business (Bender, 2007).

I collect digitized versions of these records and construct a database of 1,253,317 Scandinavian first-time emigrants over the period 1869-1920. Foreigners and visitors are identified and removed from the database. For 88% of these migrants, information on full name and gender is also available, and so is information on birthplace for 25% of these.[8] The passenger lists do not cover all emigrants of the period, but compared to official counts of emigrants described below, 70% are accounted for in the migrant database of this paper. Around 300,000 emigrated before the first data point and are not included. The database is nonetheless the most comprehensive of its kind.

To test the overall reliability of the passenger list data, I compare it to the corresponding population census records. In particular, I compare the change from one population census to the next in the number of people that share the same first name, birth decade, and residence district to the number of emigrants with the same characteristics leaving between these census years. The results in Appendix Table A.7 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.

3.2.2 Aggregate Migration Flows

Part of the empirical analysis focuses on the relationship between emigration and cultural change at the subnational district level. Rather than relying entirely on passenger list data for that purpose, I use additional official emigration statistics from Norway and Sweden. These data were constructed at the time based on church records, and they are considered a more complete source of information on *aggregate* migration flows than the passenger

[8]The main passenger list data sources are: The Copenhagen Police Emigration Protocols 1869-1908 (Denmark), "Emihamn" Emigration Records 1869-1950 (Sweden), and Norwegian regional emigration lists 1869-1930. The "Emihamn" records contain no information on birthplace. Instead two additional sources are used: "Emibas Göteborg" and "Emibas Värmland" that cover emigrants that resided in the Göteborg or Värmland districts at emigration. Information on residence and birthplace is originally stored in more or less precise strings in the source data, which I have transformed to codes identical to those used in the population census records.

list data, where a significant number of emigrants are unaccounted for.[9] Reassuringly, a comparison between the two data sources in Appendix Table A.8 reveals a high degree of within-country correlation over time.

In Denmark, where no official subnational emigration statistics exist, aggregate passenger list numbers are used instead. This solution is not flawless since the Danish passenger lists lack a few important sources of information, including Mormon emigration and documentation from travel agencies in Horsens, Vejle, and Ribe (Bender, 2007). I therefore calculate two alternative Danish district emigration variables with additional emigration sources.[10]

3.3 Cultural Proxies from First Names

Survey and experiment based indicators of cultural values do not exist historically. Instead, I construct indicators of individualism and collectivism from the distribution of first names in the population. This approach is based on research in sociology and psychology that argues that the commonness of first names can be seen as reflecting the core difference between collectivist and individualistic cultural values. While a common first name suggests a desire of the name-givers for their child to conform and fit in, an uncommon first name signals independence, originality, and the wish to stand out and differentiate oneself from the surrounding social environment.

First names are generally acknowledged to closely reflect the cultural and social preferences of the name-givers (Mateos, 2013). Across societies, giving a name is regarded as one of the most important acts surrounding the birth of a child. It is often the center of ceremonial celebration and connected to future hopes and aspirations on behalf of the child (Tan, 2004). First names are used to distinguish people from each other, but they also signal attachment to a number of groups such as gender, age, and nationality. The balance between independent and group identity has been identified as the most important driver of first name choice in the seminal work by Lieberson (2000).

That a preference for uncommonness in first name choice reflects individualistic values is not just a theoretical deduction. Studies on families in the United States show that

[9] The underlying individual level church record data is not available in any complete format and is therefore not used in this paper.

[10] One uses passenger list counts for the period 1850-1920 from ships leaving Hamburg in Germany that carried a total of 2/3 of all Danish emigrants (Bender, 2007). Another combines the Danish passenger list counts with the Hamburg data from before 1869, Mormon data from 1872-1887, and the total number of ship tickets sold by the Ribe, Vejle, and Horsens agencies. Aggregate data from the Hamburg lists are produced by searching the records at ancestry.com. The subnational distribution of emigrants from the Mormon lists were constructed by counting the number of results of searching for the names of the Danish regions. In case of the local Vejle, Ribe, and Horsens, aggregate emigration numbers were based on the assumption that most of their customers resided in the same region as the location of the ticket office. Sources are found in Bender (2007). None of the additional data sources are readily available at the individual level. Appendix Table A.9 shows that the number of Hamburg list emigrants is not particularly correlated with the passenger numbers preserved by the Danish police, which underlines the importance of checking future empirical results with both variables.

parents who choose less common names for their children motivate this by a desire for the children to be unique and different (Zweigenhaft, 1981; Emery, 2013). Bearers of uncommon names have furthermore been documented to score low on values associated with conformity (Schonberg and Murphy, 1974). Conversely, Dutch families that choose more common first names have been found to spend more time with other people of their local communities (Bloothooft and Groot, 2008).

Previous research has used aggregate trends in first name uncommonness to study cultural differences across the US and Japan (Varnum and Kitayama, 2010) and to document rising individualism over time for a number of Western countries (Lieberson and Lynn, 2003; Twenge et al., 2010).

Using first names to construct indicators of culture is associated with a lot of advantages. Cultural indicators can be constructed at any point in time and at any aggregation. Another useful feature is the fact that the child has no influence over its name. The properties of a first name reflect the cultural environment in which the child grows up and from which it learns. The uncommonness of a first name thereby measures inherited traits of the person carrying the name. When growing up, a child may evolve a different set of cultural preferences as it learns and interacts with other members of society. The predetermined nature of first names provides a useful source of identification.

The rest of this subsection is devoted to further corroboration of the link between first name commonness and individualistic and collectivist values. Details on how I calculate the commonness of first names in the historical setting follow in section 3.3.2.

3.3.1 Validation

Using Present-Day Indicators To further validate that a preference for common first names is linked to collectivism over individualism, I compare contemporary first name patterns to better known cultural indicators across and within countries. For this purpose, I collect all publically available statistics on the ten most popular baby names in 2015 across the world.[11] Since the entire distribution of first names is rarely available, I calculate the measure of first names commonness as the share of newborns with names among the ten most popular female or male names of the country or subnational district.[12]

To examine how well first name patterns predict culture across countries, I use the Hofstede (2001, 2010) index of individualism, which has been used extensively in the empirical literature (Varnum and Kitayama, 2010; Buggle, 2017; Gorodnichenko and Roland, 2011a,

[11]This is done 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). For some countries only the most popular male or female name is identified. The result is a database with national level statistics for 38 countries and subnational district statistics for 13 of them.

[12]Some countries (including Sweden) report statistics on just the most popular first name, which are added to the district level sample)

2011b). Based on surveys among IBM employees across the world, Hofstede constructs six measures of core cultural traits, where individualism is defined as a preference for a loosely-knit social framework in which individuals are expected to take care of themselves and their immediate families. Results in Appendix Table A.2 show that the commonness of baby names correlates negatively with individualism. The correlation is robust to a number of controls, including measures of diversity (ethnic, genetic, and linguistic) and other Hofstede cultural indicators. National first name patterns also correlate significantly with other more or less known indicators of individualism and collectivism as seen in Appendix Table A.3.[13]

To ensure that the commonness of first names also correlates with culture within countries, I use individual level data from the European and World Values Surveys (EVS/WVS). I calculate the Welzel (2013) measure of *'emancipative values'* for each individual based on attitudes towards lifestyle liberty, gender equality, personal autonomy, and the voice of the people.[14] This measure is recommended by the World Values Survey Association as the best indicator of individualism in the survey. I add country fixed effects to study within-country variation, control for individual respondent characteristics, and cluster standard errors at the district level. Results in Appendix Table A.4 show that individuals that live in districts with common first names are less likely to point out individualistic values and beliefs as being important to them. This is especially true in Scandinavia and when focusing on cultural values linked to personal autonomy and lifestyle liberty.

Using Historical Language Data To gauge whether change in first name commonness over time is associated with cultural change in individualism, I carry out another validation check with historical data for Sweden. The alternative indicator of individualism is now based on the language structure of 20 local newspapers that cover 12 Swedish districts[15] for the years 1780-1900. Using all digitized copies of these newspapers, I use the ratio of first person singular (I, me, mine) over plural (we, us, our) pronouns as an indicator of

[13]Including all relevant controls, this holds for a number of survey-based indicators like the Schwartz (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 & Kashima, 1998; Abdurazokzoda and Davis, 2014; Meyer-Schwarzenberger, 2015). The set of indicators is inspired by what is used in Buggle (2017).

[14]Specifically the measure is based on attitudes to homosexuality, abortion, and divorce (*lifestyle liberty*); opinions on women's right to jobs, political careers, and education (*gender equality*); the emphasis on independence and imagination over obedience as goals in child rearing (*personal autonomy*); and beliefs that people should have more say in government and local affairs (*voice of people*).

[15]The newspaper titles are Aftonbladet (Stockholm), Blekingsposten (Blekinge), Bollnäs tidning (Gävleborg), Dalpilen (Dalarna), Fahlu weckoblad (Dalarna), Faluposten (Dalarna), Göteborgs weckoblad (Göteborg och Bohus), Gotlands tidning (Gotland), Jönkopingsbladet (Jönköping), Kalmar (Kalmar), Lindesbergs allehanda (Örebro), Norra Skåne (Kristianstad), Östergötlands veckoblad (Östergötland), and Wernamo tidning (Jönköping). They are available for linguistic analysis at *https://spraakbanken.gu.se/*

individualism. The use of different pronoun classes reflects a focus on the individual over the collective (see summary by Oyserman and Lee, 2008).

Using newspaper and time fixed effects, results in Appendix Table A.5 reveal that decadal changes in the relative use of pronouns correlates significantly and negatively over time with the commonness of first names given to cohorts born in the same district and decade that the newspaper was published. This is true when the measure of first name uncommonness is based on the 10 most popular male and female names, and when it is calculated as the share of a cohort that carry the same name.

3.3.2 My measure of first name commonness

I calculate the commonness of a first name as the *share of a birth cohort that carries the same name*. Birth cohorts are defined by a birth decade, birth district, and gender to ensure both enough observations to calculate precise measures, that the measure reflects the culture of local social networks, and that changing naming fashions do not impact the measurement.[16]

A few practical challenges are worth further discussion. First, errors are present in all archival material. Names may be wrongly spelled due to illiteracy or different enumerator practices. In my baseline measure of first name uncommonness I use the original spelling, because differences in spelling may also signal a preference for uniqueness. I do however check the robustness of all empirical results by using the phonetically spelled version of first names, whereby small discrepancies in spelling (i.e. *Christian* versus *Kristian*) are eliminated. Third, some people are noted to have multiple first names (with an average of 1.5 names across all individuals in the censuses). I focus on the initial first name in the baseline measure and use the full list of first names in robustness checks.

Second, I construct a number of variables to control for aspects in first name commonness that do not reflect collectivist and individualistic values. This includes gender and sibling structures.[17] I also construct indicators to account for the tradition in some families to choose a name of an older family member. Moreover, the first name of an individual may be uncommon not because of the emphasis on individuality but because of a social or physical distance to the dominant culture of a location. Although Scandinavia as a region is considered to be extraordinarily homogenous in terms of ancestry, shared history, and interconnectedness, different aspects of diversity may still be an issue in the measurement. I capture this with a measure of last name uncommonness, which

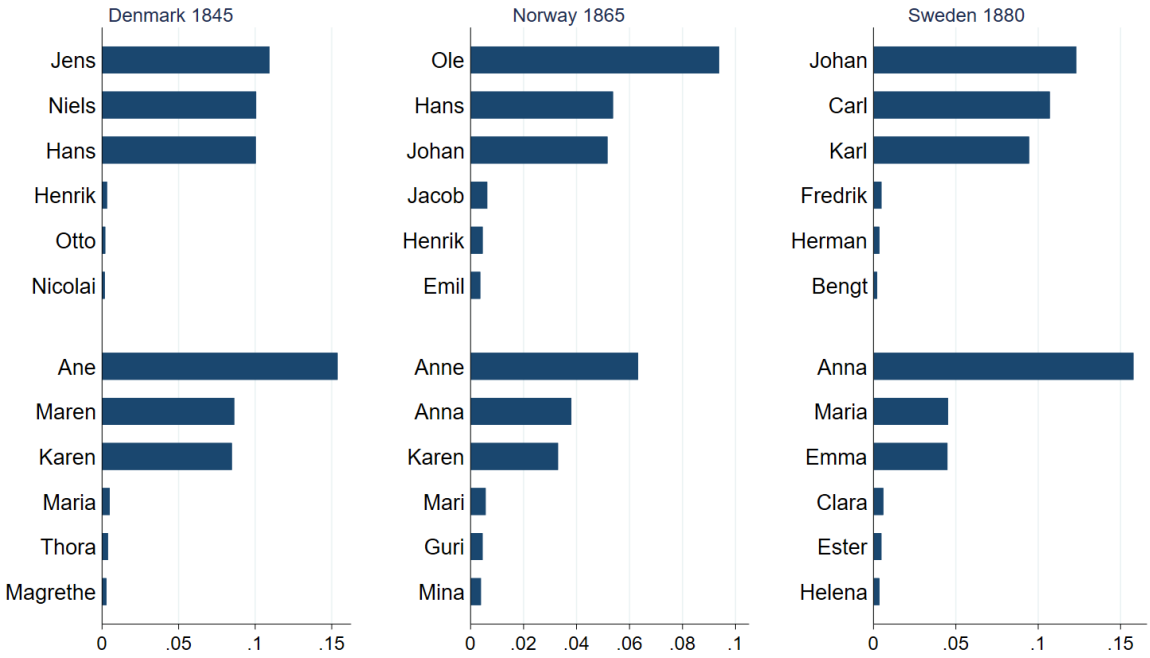
[16]Had I instead looked at national naming patterns, the uncommonness measures would also capture national connectivity and attitudes towards the state. Similarly, if I had included the entire population across birth decades, changing naming fashions would impact the measurement. Using birth decades instead of single birth years ensures enough observations to calculate precise measures.

[17]Research has documented how firstborn children and boys in general receive more common names (Lieberson 2000). The underlying intuition is that since these individuals traditionally have been assigned the role of providers in the family, their ability to blend into general society has been more necessary.

since Darwin (1875) has been used to proxy genetic, ancestral, socio-economic, or geographic diversity (Cavalli-Sforza et al., 1994).[18] Where available, other aspects of social and economic circumstances are controlled for.

Although I have just emphasized that I calculate measures of first name commonness based on local naming patterns, an illustration of a few nationally representative examples of common and uncommon first names may be helpful in understanding later empirical results. Figure 3 plots the three most common and three uncommon first names for boys and girls among children below the age of 10 in the earliest population census available for each country. It clearly shows that first name were highly concentrated at this point in time. A large share of children were given the same few names. Looking at the three most common male and female children's names across the censuses in Appendix Table A.6 reveal that although there was little replacement of first names in top, the concentration of first names became weaker in the top and more evenly distributed.

Figure 3. Examples of common and uncommon first names



Notes: The top 1, 2, 3, 30, 40, and 50 most popular names among children (<10 years old) in the earliest census for each country. The x-axis shows the share of children that carry a particular name. Female names are listed in the bottom and male names in the top.

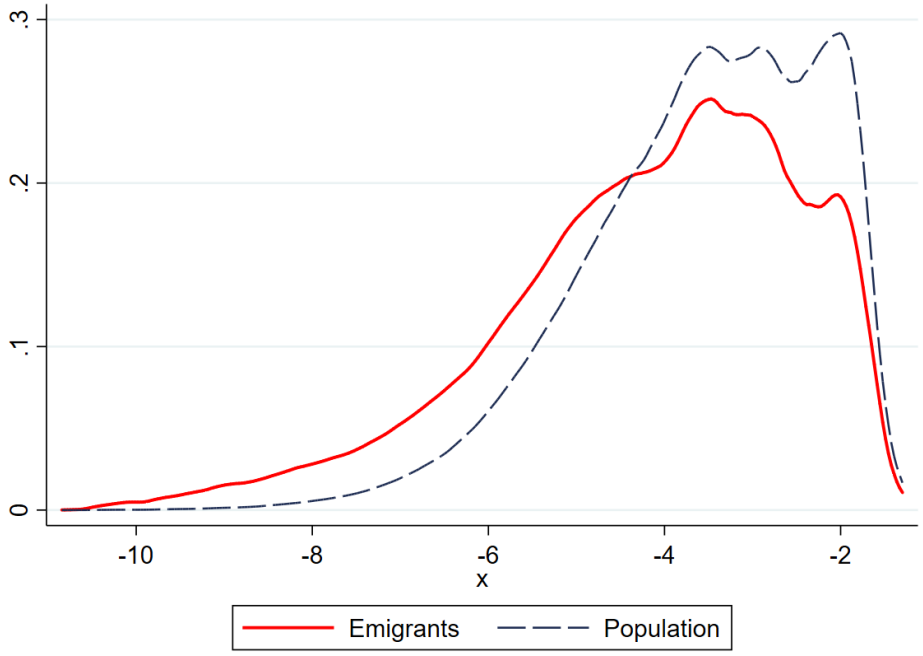
Imputing the First Name Commonness of Emigrants As noted, I define the commonness of first names relative to birth cohorts and birth districts. I calculate the commonness of emigrant first names relative to their birth cohorts with a few simple

[18] Around 35% of the populations included in the data were given last names according to patronymic practices, whereby a suffix is attached to the father's name. These were more widespread in Norway, and less so in Denmark. In these cases, the commonness of last names does not carry direct historical signals. They do however still control for a family history of individualism, which may be impacted by the mentioned confounders.

steps. For each first name carried by census residents in a certain district, birth decade, and gender, I compute the average commonness of this name relative to the birth districts of the people carrying this name. I then assign these probabilistic measures of commonness to each emigrant based on the population census closest to birth year. Emigrants that migrated before the census year, are added to the census before calculating the first name uncommonness.

Figure 5 shows the distribution in first name commonness (in logs) across all individuals that lived during the Age of Mass Migration. The dashed line represents that of the population and the red solid line that among emigrants. Here it is clear that emigrants had more uncommon first names than the general population. More sophisticated analysis is necessary to test the robustness of this relationship.

Figure 4. Distribution of first name commonness



Notes: First name commonness is calculated as the share of birth cohorts (by birth decade, birth district, and gender) that share the same first name. Here shown in logs. Emigrants are individuals from the passenger lists. The population includes individuals that lived during the period, including both emigrants and stayers. I ignore population cohorts with no emigration.

4 Selective Emigration

4.1 Linked Sample

To test if emigrants selected on cultural traits, I link emigrants and stayers to their childhood households. First, I create a list of all males who towards the end of the Age of Mass Migration had either emigrated or stayed in Scandinavia. Stayers are identified from the most recent population census from each country: The 1901 census in Denmark

and the 1910 census in Norway and Sweden. Emigrants are identified from the passenger lists.[19] Second, I create a subsample of male emigrants and stayers who were below the age of 15 in any of the earlier censuses. Third, I link this subsample to their younger selves based on their first names, birth year, and their district of origin. Females are left out of the sample due to the custom of surname change at marriage, which makes them difficult to link across sources.

My linking procedure is inspired by what has been standard in the related literature (Abramitzky et al., 2012, 2013; Long and Ferrie, 2013; Fouka, 2019). I identify potential links with the phonetic spelling of first and last name, the district of origin, and a two-year band around the year of birth.[20] Then follows an iterative process, where I evaluate the quality of these links based on the string distance between first and last names, using the Jaro-Winkler algorithm by Mill (2012), and birth year accuracy (see Appendix B for further details). This leaves me with a linked sample of 936,125 children that represent 36% of the total child population.[21]

I link emigrants and stayers *jointly* to minimize potential biases from the linking process. The alternative would have been to link only emigrants and then treat unlinked children as stayers. The problem with the alternative strategy is that it is easier to link people with less common names. I would not be able to compare my main cultural proxy, the commonness of first names, across linked and unlinked observations. Applying the same linking process to both emigrants and stayers eases the comparison between these groups in this regard.[22] To further eliminate concerns on this matter, I do robustness checks on the entire sample of emigrants and stayers in subsection 4.3.2.

Table 1 reports summary statistics for the linked sample and balance tests compared to the total child population. As expected, I am more likely to link individuals with uncommon names. Children from households with urban residence, lower birth order, and higher socioeconomic status are over-represented in the linked sample. I control for these characteristics in the empirical analysis. Given the sample sizes, there are significant differences in all other variables too, but they are economically small.

[19] Only emigrants with emigration year before the latest census are included.

[20] I have limited information on emigrant birth places and use their latest district of residence instead. If I am unable to locate them in their latest residence district, I look for them in the country as a whole.

[21] From the subsample of emigrants and stayers towards the end of the Age of Mass Migration, this sample represent 47%.

[22] More than half of the unlinked children remain so due to having multiple links: They share their name, birth year, and birth (or residence) district with other children or more than one individual in the combined passenger lists and most recent population census. Other sources of error are missing, insufficient, and wrong personal details. This is more likely with emigrants in the passenger lists since they are less complete and only record the latest district of residence. While this location generally reflects the latest permanent residence (Karadja and Prawitz, 2019), emigrants may still have moved away from their childhood districts. This makes them more difficult to link.

Table 1. Summary statistics and balance tests

	Linked sample		Population		Difference	
	Mean	S.D.	Mean	S.D.		
Main variables:						
Emigration	0.054	0.227				
Out-migration	0.170	0.376				
First name cmn.	0.034	0.045	0.048	0.053	-0.014	(0.00006)
Last name cmn.	0.039	0.051	0.051	0.053	-0.012	(0.00006)
Log first name cmn.	-4.502	1.838	-4.027	1.827	-0.475	(0.002)
Log last name cmn.	-4.704	2.259	-4.148	2.153	-0.556	(0.003)
Age	7.335	4.287	7.594	4.360	-0.259	(0.005)
Brothers	1.634	1.409	1.526	1.364	0.108	(0.002)
Birth order	1.903	1.214	1.800	1.148	0.102	(0.001)
Official name	0.931	0.254	0.951	0.216	-0.020	(0.0003)
Non-rare name	0.986	0.118	0.988	0.109	-0.002	(0.0001)
No. of names	1.653	0.623	1.629	0.616	0.024	(0.0007)
Length of name	5.138	1.305	5.064	1.308	0.074	(0.002)
Name shared by family member	0.205	0.404	0.244	0.430	-0.039	(0.0005)
Family size	6.250	2.642	6.062	2.766	0.188	(0.003)
Related to hh. head	0.986	0.117	0.967	0.178	0.019	(0.0002)
Urban residence	0.189	0.392	0.167	0.373	0.023	(0.0005)
Farm	0.451	0.498	0.436	0.496	0.015	(0.0006)
Servants in hh	0.179	0.383	0.190	0.392	-0.011	(0.0005)
Father characteristics:						
Poor relief	0.008	0.088	0.011	0.106	-0.004	(0.0001)
Assets	0.267	0.443	0.231	0.421	0.037	(0.0005)
Labor force	0.855	0.352	0.799	0.400	0.056	(0.0005)
Non-manual work	0.080	0.271	0.060	0.238	0.020	(0.0003)
Skilled work	0.441	0.496	0.392	0.488	0.049	(0.0006)
Foreigner	0.020	0.140	0.014	0.118	0.006	(0.0001)
Additional proxies:						
Log old sibling first name cmn.	-4.286	1.860	-3.882	1.822	-0.404	(0.002)
Log past decade first name cmn.	0.023	0.043	0.038	0.054	-0.015	(0.00006)
Live in parent's birthplace	0.820	0.384	0.822	0.382	-0.002	(0.0005)
Parents same birthplace	0.778	0.416	0.782	0.413	-0.004	(0.0005)
Live with grandparents	0.063	0.242	0.065	0.247	-0.003	(0.0003)
Live with other relatives	0.160	0.367	0.159	0.366	0.001	(0.0004)

Notes: Linking conducted by first name, last name, age, and district of residence. 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=936,125; Population N=2,599,846.

4.2 Empirical Model

Using the linked sample of emigrants and stayers, I estimate the following model:

$$M_{inc} = \alpha_p p_{nc} + \phi_c + \alpha_X X_{inc} + \varepsilon_{inc}, \tag{1}$$

where M_{icn} indicates the emigration of individual male i with first name n and birth cohort c (defined by birth decade and birth district), who was below the age of 15 in an early population census. It equals one if the individual is observed in the passenger lists, and zero if he is observed in the most recent population census for each country. $p_{nc} = ln(\frac{N_{nc}}{N_c})$ is the commonness of his first name, which proxies having been raised in a collectivist household. ϕ_c denote cohort fixed effects, and X_{inc} is a set of control variables of the individual child, his first name, and his childhood household.

A negative α_p implies that individualistic self-selection into emigration takes place. Identification relies on the pre-determinedness of explanatory variables and the ability to hold other relevant childhood aspects fixed.

4.3 Main Results

Table 2 reports results from estimating Model 1 on the linked sample of emigrants and stayers. All regressions control for age in childhood, number of siblings, birth order fixed effects, and cohort fixed effects. Standard errors are clustered at the birth district level.

In columns (1) to (3) the full sample of linked emigrants and stayers is considered. In column (1) only baseline controls are included. Column (2) controls for last name commonness and an indicator of carrying a first name that is phonetically identical to that of an older family member. The first is included to capture other aspects of social similarity and distinction, and the latter to proxy family traditions. I also control for name string length and number of first names here. In column (3) household characteristics related to socioeconomic status and occupation are controlled for. This includes dummies for urban and farm residence, having servants employed, being related to the household head (to proxy social status within the household), and family size. Separately for the mother and father, I control for them being on poor relief, owning assets (land, shops, or factories), being in the labor force, having skilled and non-manual work (calculated from reported HISCLASS codes), and being born in another country.

The estimated coefficient on collectivist first names is negative in these specifications, which is in line with the hypothesis that children from more collectivist household are less likely to emigrate. Comparing the coefficients to the summary statistics in Table 1, a one standard deviation change in first name commonness is associated with a $0.06 - 0.07$ standard deviation change in the propensity to emigrate.

In column (5) I exclude children that emigrated before the age of 15 in order to rule out that parents gave their children less common names while planning their own emigration. Column (6) excludes children with mothers above the age of 43. Similarly to Abramitzky et al. (2012) this is to ensure accurate sibling and birth order measures. In columns (7) and (8) I analyze the decision to emigrate and out-migrate separately. Out-migrants are defined as children that in the most recent census lived outside their district of birth. While there is a negative selection of people with common first names into internal migration, the coefficient is smaller than with emigration. This should be expected as moving abroad likely entails larger social costs.

Table 2. Collectivist cultural traits and emigration

Dep. variable	(1) Emigration	(2) Emigration	(3) Emigration	(4) Emigration	(5) Emigration	(6) Emigration	(7) Out-migration
First name cmn.	-0.00728***	-0.00899***	-0.00879***	-0.00798***	-0.00733***	-0.01012***	-0.00300***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Excluding				Child emigrants	Mothers above 43 yrs.	Out-migrants	Emigrants
Controls:							
Individual	X	X	X	X	X	X	X
Name char.		X	X	X	X	X	X
Household char.			X	X	X	X	X
Mean (sd) of dep. var.	0.054 (0.22)	0.054 (0.22)	0.054 (0.22)	0.050 (0.22)	0.045 (0.21)	0.065 (0.25)	0.17 (0.38)
Observations	936,125	936,125	936,125	931,371	577,767	785,756	885,267
R-squared	0.17	0.17	0.18	0.17	0.18	0.23	0.09

Notes: OLS results. The sample consists of males below the age of 15 in the earliest censuses (DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, SE-1900). In columns (1)-(6) the emigration indicator equals one if the individual is observed in the passenger lists and zero if observed the later censuses (DK-1901, NO-1910, SE-1910). In column (7) out-migrants are identified as living outside one's birth district the later censuses. All regressions include cohort (birth decade X district) fixed effects. Individual controls include age in the census, number of siblings, and birth order fixed effects. Additional first name characteristics are string length, number of first names, and dummies for rare names, officially recognized names, and for sharing name with a family member. Household controls include urban location, being engaged in farming, having servants employed, family size, and separately for the mother and father: being on poor relief, owning assets, being in the labor force, having skilled work, having non-manual work, and being foreign. Standard errors clustered at the district level are shown in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The main results are robust to using alternative measures of commonness as seen in Appendix Table A.10. This includes a dummy for the name being among the 10 most popular in the cohort, and commonness measures based at the country or municipality level. Different kinds of spelling are also considered, including the phonetic equivalent and the first single and three letters of the name. Rare names that are shared by less than 10 people in a census may be dropped, and so may names that do not appear on official names lists today.[23] The robustness of the results indicate that they are not driven by size of the reference group, spelling mistakes or alternative spellings of the same name. The results are also robust to additional fixed effects for birth municipality, last name, father's occupation, and father's ownership of assets, as seen in Appendix Table A.11.

4.3.1 Group-Level Results

The biggest concern with the results above is that they are driven by biases in the linking process. I therefore carry out a group-level analysis that includes all emigrants in the passenger lists, male and female. I link them to groups in the population census closest to birth year that share their first name, gender, birth decade, and district of residence. I estimate the model:

$$m_{nc} = \alpha_p p_{nc} + \phi_c + \alpha_X X_{nc} + \varepsilon_{nc}, \tag{2}$$

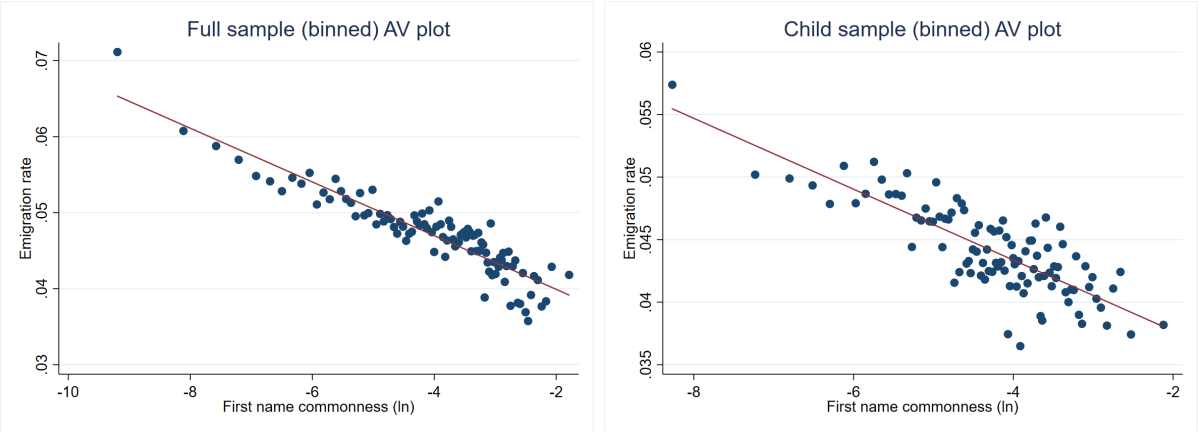
where $m_{ncdy} = \frac{M_{nc}}{N_{nc}}$ is the total emigration rate (1870-1920) among people in cohort c (defined by birth decade and residence district) with first name n. This is regressed on the commonness of the first name, $p_{nc} = ln(\frac{N_{nc}}{N_c})$, cohort fixed effects, ϕ_c, and a vector of control variables, X_{ncdy} (averaged over the group that share the same first name and cohort).[24] The model represents a version of Model that is collapsed over individuals that share the same first name and cohort. Control variables are less precisely measured, but they capture if some first names are particular to certain types of people.

Regressions largely replicate those in Table 2, and results are shown in Appendix Table A.12. AV plots in Figure 5 shows the AV plots of the full specification on the entire emigrant sample (left) and on emigrants that were below the age of 15 in any given census (right). Again, common first names significantly predicts who chooses to stay rather than emigrate, and this is robust to a host of relevant factors that are characteristics of the people who carry the same name.

[23] Each Scandinavian country has a list of officially approved names today, which I use in this test.

[24] Note that cohorts here are defined by birth decade, gender, and residence district (instead of birth district). The commonness of a first name in a cohort is calculated as the over individuals that share this name and residence cohort but have different birth districts. For 4% of the emigrants, I cannot find a group with the same name in the census, and they are dropped from the analysis.

Figure 5. Group Level AV-plots



Notes: AV-plots from regressions where the unit of observation is a group that shares the same first name, birth decade, gender, and district of residence. The dependent variable is the share of this group that emigrates over the Age of Mass Migration, and the independent variable is the commonness of the first name relative to the cohort. The plot to the left considers all cohorts and includes cohort fixed effects and additional first name characteristics as controls. The plot to the right considers only cohorts that were below the age of 15 in a given census. Average sibling and household controls are included here as additional controls.

4.3.2 Alternative Cultural Proxies

Table 3 reports results from estimating the main specification from model (1) with alternative cultural proxies. In column (1), I consider the first name commonness of the oldest brother in the household. Complementing my use of birth order fixed effects in all regressions, this is to alleviate concerns that effects of collectivist names on emigration is driven by birth order. The negative and significant coefficient suggests that the estimated relationships between first name commonness and emigration largely captures cross-family differences. Another concern is that parent lack foresight and imagination and may choose first names that are revealed to be more common than intended. In column (2), I therefore use the commonness of the child's first name based on the distribution in the decade before the child's decade of birth. This concern does not appear to be what is driving the main results.

Other characteristics of collectivist households are that they tend to stay connected to their local neighborhoods and live in extended family structures. In columns (3) and (4) I include proxies of the former, and in columns (5) and (6) of the latter. Results reveal a lower propensity to emigrate among children who lived in the birth district of either parent, whose parents were born in the same district, who lived with a grandparent or other distant relatives (defined by not being a member of the nuclear family).

Table 3. Collectivist cultural traits and emigration, using alternative cultural proxies

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable	Emigration					
Log first name cmn. (oldest brother)	-0.0041***					
	(0.000)					
Log first name cmn. (past decade)		-0.0059***				
		(0.001)				
Live in parent's birthplace			-0.0084***			
			(0.001)			
Parents same birthplace				-0.0053***		
				(0.001)		
Live with grandparents					-0.0051***	
					(0.001)	
Live with other relatives						-0.0025**
						(0.001)
Controls	X	X	X	X	X	X
Mean (sd) of dep. var.	0.054 (0.22)	0.054 (0.22)	0.054 (0.22)	0.054 (0.22)	0.054 (0.22)	0.054 (0.22)
Observations	936,125	936,125	936,125	936,125	936,125	936,125
R-squared	0.17	0.17	0.17	0.17	0.17	0.17

Notes: OLS results. The sample consists of males below the age of 15 in the earliest censuses (DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, SE-1900). The emigration indicator equals one if the individual is observed in the passenger lists and zero if observed the later censuses (DK-1901, NO-1910, SE-1910). Controls include age in census, number of siblings, birth order fixed effects, cohort fixed effects, last name commonness, additional first name characteristics (string length, number of names, rare names, officially recognized names, and name shared with a family member), and household controls (urban, farming, employing servants, family size, and for the mother and father: being on poor relief, owning assets, being in the labor force, having skilled work, having non-manual work, and being foreign). Standard errors clustered at the district level are shown in parentheses.. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.4 Additional Results

4.4.1 Determinants of Selection Intensity

Not all emigrants in my data are equally selected. Some have more common first names and some more uncommon first names. Here I examine the source of this variation. I create a 10-year panel based on my sample of linked emigrants and stayers over the period of the Age of Mass Migration (1850-1920), and I code their emigrant status with information on the year of emigration in the passenger lists. This means that emigrants are only coded as so in and after their decade of emigration. With this data, I estimate a version of Model 1 where the dependent variable now includes a time subscript t, and where I interact time-invariant explanatory and control variables with time-varying local circumstances, l_t:

$$M_{inct} = \alpha_p p_{nc} \times l_t + \phi_c + \phi_t + \alpha_X X_{inct} + \varepsilon_{inct}, \tag{3}$$

I estimate the model with all control variables from Table 2 and add time fixed effects. Results are shown in Table 4, where, for simplicity, coefficients on the time-varying factors are unreported. In columns (1) and (2), I interact the indicator of collectivist first names with the rate of emigration that took place in the same district and decade. When a larger share of an individual's residence population emigrated, the effect on emigration of one's first name was weaker. The same is true in periods when the total rate of past emigration was higher, as seen in columns (3) and (4). This is in accordance with the idea that a constraint on the emigration of people with collectivist traits is the lack of familiar social networks abroad.

In columns (5) and (6) I explore the role of weather shocks, using the intensity by which temperatures dropped below their long-run averages. Temperature drops were particularly harmful to Scandinavian agriculture and pushed people into emigration (Karadja and Prawitz, 2019).[25] The results show that selection on collectivist traits weakened in decades with more weather shocks. Weather could have lowered the relative social costs of emigration through two channels. First by pushing more people into emigration, whereby migrant networks abroad would grow. Second, by increasing the economic returns to migration.

[25]Karadja and Prawitz (2019) find that local frost shocks in 1864-1867 predicted emigration outflows from Swedish municipalities. I construct a similar, but less detailed, measure of temperature drops that cover all of Scandinavia. The underlying data used in Karadja and Prawitz (2019) cannot be obtained for all of Scandinavia. Instead I use data from Luterbacher, Dietrich, Xoplaki, Grosjean and Wanner (2004), which was also used in Sequeira et al. (2019). For each year, season (spring, summer, fall, winter) and 0.5 degree pixel, this data reports the mean temperature. For each pixel and decade, I calculate the number of seasons where the temperature dropped more than one standard deviation below the long-term mean. I then average this across pixels within the subnational district.

Table 4. Factors that impact selection (panel estimates)

	(1)	(2)	(3)
Dep. var.		Emigration	
Log first name cmn.	-0.0103***	-0.0097***	-0.0137***
	(0.000)	(0.000)	(0.000)
Log first name cmn. X moderator var.	0.0429***	0.0079***	0.0011***
	(0.006)	(0.001)	(0.000)
Moderator variable	District emigration	Past district emigration	Frost shocks
Mean(s.d.) dep. var.	0.060 (0.24)	0.060 (0.24)	0.060 (0.24)
Observations	1,750,285	1,750,285	1,750,285
R-squared	0.14	0.14	0.14

Notes: OLS results from a panel of males below the age of 15 in the earliest censuses (DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, SE-1900), observed in each decade 1850-1920 in which they were alive. For emigrants, the emigration indicator equals one in and after the decade of emigration and zero before emigration. For non-emigrants it equals zero in all decades. First name commonness is the log share of one's birth cohort that share the same first name. This is interacted with district level emigration rates of the same decade in columns (1) and (2), total past emigration in columns (3) and (4), and the number of frost shocks occurring in the decade in columns (5) and (6). Regressions control for cohort, decade, and birth order fixed effects, mid-decade age and age squared of each individual, and the full set of control variables from Table 2. Standard errors are clustered at the district level. Significance levels: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

4.4.2 Evidence on the Behavior of Emigrants

The results above indicated that people with collectivist traits emigrated when the social costs of doing so were relatively lower. Here I directly test if cultural traits mattered for *how* people emigrated. I expect that collectivist people emigrated collectively, with others from their social networks.

First, I look at three parameters of emigrant behavior observed in the passenger lists. For a subsample of the Norwegian emigrants, I know their stated motivation for emigration. This information was asked and recorded as part of the process of buying a ticket for the trans-Atlantic ships. The most frequent motivation was employment and higher wages, but a 24% stated that they left to join someone, often a family member, abroad. For a subsample of all emigrants, I know if they emigrated with family or not. Finally, I know if people emigrated on the same date as other people from their hometowns. Table 5 presents OLS results from regressing indicators of collective emigration on the commonness emigrants first names. Controls included are age at emigration, last name commonness, cohort fixed effects, and other first name characteristics. In every other

column, I control for the overall likelihood of emigration among people that share the same first name and cohort.[26] The results reveal that emigrants with common first names were more likely emigrate collectively. They were more likely to join someone abroad, emigrate with family, and emigrate with fellow townspeople. The results support the idea that people with common names value strong social networks, and that this impacts their emigration behavior.

Table 5. Collectivist cultural traits and emigration style

Dep. var.	(1) Emigrate with purpose of meeting someone	(2) Emigrate with purpose of meeting someone	(3) Emigrate with family	(4) Emigrate with family	(5) Emigrate with fellow towns people	(6) Emigrate with fellow towns people
Log first name cmn.	0.0089***	0.0106***	0.0090***	0.0094***	0.0048***	0.0059***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Pr(emigration) control		X		X		X
Mean (sd) of dep. var.	0.24 (0.43)	0.24 (0.43)	0.38 (0.49)	0.38 (0.49)	0.74 (0.43)	0.74 (0.43)
Observations	31,521	31,521	441,776	441,776	758,638	758,638
R-squared	0.47	0.47	0.22	0.22	0.10	0.10

Notes: The sample consists of all emigrants observed in the passenger lists. Results are from OLS regressions of emigration style (collective versus independent) on the commonness of the emigrant's first name relative to his or her birth cohort. All regressions include cohort (birth decade, gender, last district of residence) fixed effects and control for age and age squared in the year of emigration, the log commnoness of last names, and first name string length. Columns (2), (4), and (6) control for the probability of emigration, calculated as emigrant the share of people in the population that share the same first name and cohort. Standard errors clustered at the last residence district level are shown in the parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Second, I look at the behavior of Scandinavian emigrants in the US. Here I expect collectivist emigrants to settle and interact with fellow Scandinavians. With data from the 1900 and 1910 population censuses in the United States, I show in Appendix Table A.13 that Scandinavian immigrants with more common names relative to their home countries were more likely to (after arrival to the US) marry someone of own nationality, more likely to settle in states with higher concentrations of people of own nationality, and more likely to choose Scandinavian sounding first names for children born in the US.[27,28]

[26] These numbers are taken from the data in section 4.3.1.

[27] Swedish, Norwegian, and Danish born migrants are identified in the 1900 5% and 1910 1% NAPP samples for the United States, which also provide information on the year of immigration. I calculate the commonness of their first names with the Swedish or Norwegian 1900 or the Danish 1880 population census. Since I do not know their district of origin, first name commonness represent national measures. The set of control variables used in the regressions is similar to but different from the rest of the empirical analysis. All regressions control for years spent in the US, which as expected is associated with stronger assimilation outcomes.

[28] I calculate the Scandinavian distinctiveness of first names according to the method in Fryer & Levitt (2004). The Swedish, Norwegian, or Danish distinctiveness of a first name is not directly comparable to its commonness. A name that was common in Scandinavia may not be especially distinctive in the US.

These results corroborate my findings on selective migration from Scandinavia as they confirm expected emigrant behavior once reaching destination.

5 Emigration and Cultural Change

5.1 Cultural Shocks of Emigration

The previous section provided evidence that Scandinavians with collectivist cultural characteristics were less likely to emigrate during the Age of Mass Migration. In this section, I study if selective emigration was associated with lasting cultural change in migrant-sending districts. To do this, I need measures of the relative magnitude by which emigration altered the distribution of cultural traits in the sending-populations. More precisely, I need measures of the relative increase in sending-population collectivism caused by the departure of particularly individualistically minded emigrants. I refer to this as the cultural shock of emigration, and for any given location and time period, it can be calculated as:

$$\frac{p_{t+1} - p_t}{p_t} = \frac{m_t}{1 - m_t} * \frac{p_t - p_t^m}{p_t} \ , \tag{4}$$

where $p_t = \frac{P_t}{N_t}$ is the pre-emigration prevalence of collectivism in the population, and p_{t+1} the post-emigration level. This is a function of the rate of emigration, $m_t = \frac{M_t}{N_t}$, and the cultural distance between population and emigrants, $(p_t - p_t^m)$, where $p_t^m = \frac{P_t^m}{M_t}$ is emigrant collectivism.[29] Note that the pre-emigration population includes both emigrants and stayers, and that the equation is only useful in a computational exercise where m_t, p_t, and p_t^m are known.

I approximate relative emigrant collectivism, $(p_t - p_t^m)/p_t$, with the difference in first name commonness between emigrants and their cohorts in Scandinavia. Specifically, I calculate the cultural distance of each emigrant in the passenger lists to the cohort that shares the same birth decade, residence district, and gender in the population census closest to birth year.[30] Averaging this cultural distance over emigrants produces a measure of relative emigrant collectivism that is unaffected by the age or gender structure of emigrants and population. I then calculate the cultural shock of emigration using additional

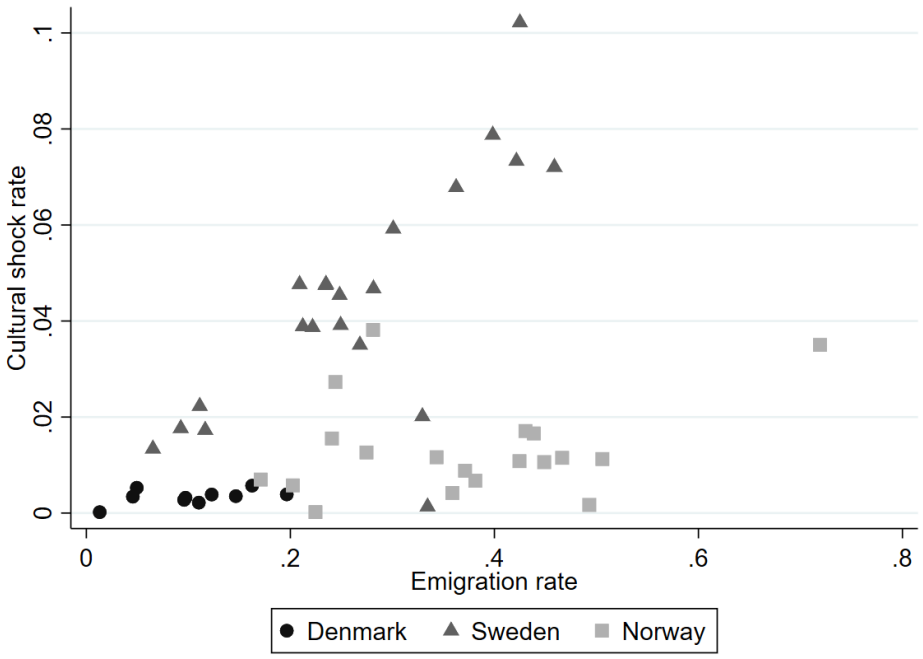
[29] To see this, consider a population, N_t, that consists of a share, m_t, that eventually migrates and a share, $r_t = 1 - m_t$, that remains. Before emigration, population collectivism is a weighted average of the two groups, $p_t = m_t p_t^m + (1 - m_t)\, p_t^r$, and after emigration it is just that of the remaining population, p_t^r. The difference between the two equals: $p_{t+1} - p_t = p_t^r - p_t = \frac{1}{(1-m_t)}\,(p_t - m_t p_t^m) - p_t = \frac{1}{(1-m_t)}\,(p_t - m_t p_t^m - (1 - m_t)\, p_t) = \frac{1}{(1-m_t)}\,(m_t p_t - m p^m) = \frac{m_t}{(1-m_t)}\,(p_t - p_t^m)$, which I then divide by p_t.

[30] As highlighted before, only the last district of residence is observed for emigrants. I calculate the commonness of first names at the residence district, birth decade, and gender level as an average of the *true* first name commonness (in relation to the birth cohort) over the residence population. Emigrants that migrated before the census year, or do not share the first name with someone in the census, are added to names distribution before calculating the commonness of first names.

information on district level emigration rates (see section 3.2.2). The shock measures are available from 1860 onwards, since these are the decades covered by the emigrant passenger lists.

Figure 6 plots the accumulated rate and cultural shock of emigration over the decades of the Age of Mass Migration for all Scandinavian districts. According to these numbers, Sweden experienced the strongest push towards increased collectivism at rates up to 10%. The cultural shocks were more modest in Norway and weak in Denmark.

Figure 6. Rate and Cultural Shock of Emigration, 1850-1920



Notes: Scatterplot of the total rate of emigration and the relative reduction in first name commonness associated with selective emigration over the Age of Mass Migration, across 50 Scandinavian districts. The reduction in home population individualism is calculated with information on total emigration and the intensity of selective emigration.

Needless to say, the cultural shocks of emigration will be measured with noise. For comparison, I calculate three other measures, which are illustrated in Appendix Figure 6. The first is based on the phonetic spellings of first names, the second uses the distribution of first names in the census closest to emigration year (rather than closest to birth year), and the third includes only emigrants and population whose first names are found in both census and passenger list records. The first two measures are very similar in level and correlate strongly with the baseline measure. The last measure is almost double in size and correlate more weakly with the baseline, with a coefficient of 0.6.

5.2 Predicting Actual Cultural Change

For emigration to generate lasting cultural change, mechanisms of intergenerational cultural transmission need to be sufficiently efficient. To empirically examine if this is the

case, I estimate the following model on the relationship between actual and emigration-induced cultural change:

$$p_{dt+1} = \rho_1 p_{dt} + \rho_2 cs_{dt} + \rho_X X_{dt} + \varepsilon_{dt} \ , \tag{5}$$

where p_{dt} is the prevalence of collectivist cultural traits in district d in the earliest census year, time t, for each country (1845 in Denmark, 1880 in Sweden, and 1865 in Norway). This is proxied by the average commonness of first names carried by children below the age of ten.[31] cs_{dt} is the sum of cultural emigration shocks that occurred between year t and 1920, calculated according to equation 4. X_{dt} is a set of control variables.

If culture is transmitted across generations, p_{dt} and cs_{dt} should predict cross-district cultural differences towards the end and after the Age of Mass Migration. I use three dependent variables to test if this is the case. The first is the commonness of children's first names in the most recent censuses (1901 in Denmark and 1910 in Sweden and Norway), the second is the commonness of baby names in 2015, and the third a collection of individual self-reported beliefs in individualistic versus collectivistic values from the 1981-2012 pooled World Values Survey (WVS) and European Values Study (EVS).

Table 6 reports the results of this estimation. All regressions control for country fixed effects, initial population levels, and basic characteristics of the children whose first names are used to proxy initial collectivist traits (gender ratio, average number of first names, and first name string length). In columns (2), (4), and (6) three additional controls are included. One is the total rate of emigration over the period to ensure that it is in fact via its initial compositional impact, and not its magnitude alone, that emigration matters for actual cultural change. The other two are the commonness of last names among the children in the early censuses, and the shock of emigration to last name commonness. As in previous analyses, this is to capture dimensions of diversity that are not related to collectivist and individualistic cultural traits.

In columns (1) and (2) the dependent variable is average first name commonness among children in the 1901/1910 censuses, and in columns (3) and (4) it is the commonness of baby names in 2015.[32] In columns (5) and (6) the dependent variable is the Welzel (2013) measure of "Emancipative Values", which is recommended as the best indicator of individualism in the WVS and EVS. The regressions here are estimated at the individual respondent level, including individual characteristics (age, age squared, gender), survey year fixed effects, and standard errors clustered at the district level. Here, negative coefficients indicate that respondent who live in districts with a higher historical concentration of first names and a stronger outflow of people with uncommon names, are less likely to

[31] Using the first names of children yields more precise measures, because measurement error caused by mortality and emigration between birth and census year is limited.

[32] First name commonness in 1901/1910 is measured as the average commonness across children below the age of 10 in the corresponding population censuses. In 2015, it is measures as the share of newborns that received one of the ten most popular boys or girls names in the district.

emphasize individualistic values as being important to them.

First, results indicate that cross-district cultural persisted over time both towards and after the end of the Age of Mass Migration. The implied standardized coefficients weaken over time, ranging from 0.7 in the medium run to around 0.15 in the long run, disregarding the insignificant coefficient in column (6). Second, and more interestingly, the compositional shocks accumulated over the Age of Mass Migration appear to have pushed cultural evolution in a collectivist direction. Again, the implied standardized coefficients weaken over time, ranging from 0.09 in the medium run and 0.05 in the long run.

Table 6. Cultural change in the medium and long run

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable	First name commonness (districts)				Individualism (survey)	
	1901/1910		2015		1981-2012	
Historical first name cmn.	0.662***	0.661***	0.277**	0.273**	-0.067**	-0.052
	(0.095)	(0.102)	(0.104)	(0.107)	(0.030)	(0.034)
Cumulative shock	1.520	2.728*	3.105***	2.767*	-0.600***	-0.638**
	(1.048)	(1.348)	(1.071)	(1.459)	(0.173)	(0.244)
Additional controls		X		X		X
Mean (sd) dep. var.	-3.8 (0.48)	-3.8 (0.48)	-2.9 (0.96)	-2.9 (0.96)	0.55 (0.17)	0.55 (0.17)
Observations	50	50	50	50	6,303	6,303
R-squared	0.96	0.96	0.99	0.99	0.34	0.34

Notes: Cross-district OLS regressions, predicting cultural differences towards the end of the Age of Mass Migration, columns (1) and (2), and today, columns (3)-(6). Historical first name commonness is the average popularity of first names given to children (< 10 years old) in the earliest census for each country (DK-1845, NO-1865, SE-1880), calculated at the district level. The cumulative shock equals the increase in first name commonness directly associated with selective emigration (see main text for explanation). All regressions control for country fixed effects, initial population levels, and basic characteristics of the children whose first names are used to proxy initial collectivist traits (gender ratio, average number of first names, and first name string length). Additional controls include the total rate of emigration over the period, the commonness of last names among the children in the early censuses, and the shock of emigration to last name commonness. In columns (1) and (2) the dependent variable is average first name commonness among children in the 1901/1910 censuses, and in columns (3) and (4) it is the commonness of baby names in 2015 In columns (5) and (6) the dependent variable is the Welzel (2013) measure of "Emancipative Values" in the WVS and EVS. The regressions here are estimated at the individual respondent level, including individual characteristics (age, age squared, gender), survey year fixed effects, and standard errors clustered at the district level. Standard errors are shown in the parentheses with the following significance levels: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Robustness I check the robustness of results in Appendix Tables A.14 and A.15. First, I divide the Welzel (2013) indicator of individualism into its four sub-components and find that the results are particularly strong in measures of personal autonomy and the lifestyle liberty values.

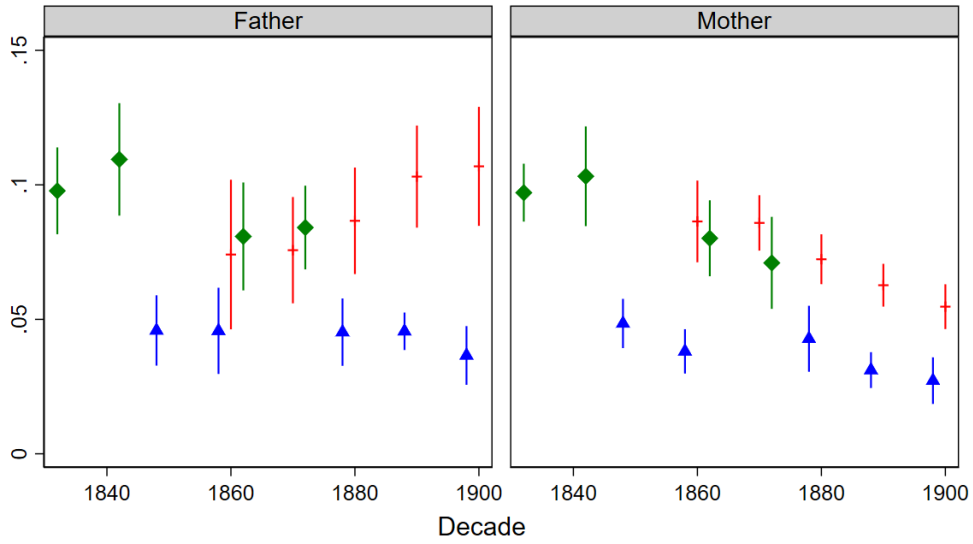
Second, I repeat the regressions from Table 6 on the pooled set of population censuses. While the above considered only the earliest census for each country, a subnational district is now represented with multiple historical values of individualism – one for each population census. Finally, I check if the results hold with alternative cumulative shock measures. This includes measures based on the phonetic spellings of first names, the

distribution of first names in the census closest to emigration year (rather than closest to birth year), including only emigrants and population whose first names are found in both census and passenger list records, and alternative sources of Danish emigration rates. Despite the small sample size, the baseline results appear remarkably robust to these alternative specifications.

Mechanism: Intergenerational Transmission As discussed in section 2.2, selective emigration can only generate lasting cultural change if mechanisms of cultural transmission are strong enough. Although this was implied by the results above, more direct evidence can be produced of the underlying mechanism.

Figure 7 illustrates significant correlations in first name commonness between parents and children. The underlying regressions consider all first-born children in the population censuses and relate their first name commonness with that of their parents and a set of relevant control variables. With elasticities between 0.05 and 0.1, the results indicate a significant transmission of conformist name giving from parents to children.

Figure 7. Parent-to-child transmission of first name commonness



Notes: Correlation in first name commonness between parents and children with birth in different decades (x-axis) in Denmark (green), Norway (blue) and Sweden (red). Coefficients on the y-axis reflect elasticities. The sample is restricted to first-borns with a mother below the age of 43. Regressions control for gender, age, and cohort fixed effects for the child and both parents. Standard errors are clustered at the child cohort level.

6 Conclusion

This paper examines the cultural determinants and consequences of the Age of Mass Migration (1850-1920) in Scandinavia. During this time period, 25% of the average populations emigrated. I collect data on nearly all emigrants and stayers from historical census records and trans-Atlantic passenger lists.

Exploiting within-cohort variation, I show that emigrants selected on individualistic cultural traits. Contrary to the emigrants, those who stayed in Scandinavia had more common first names, came from extended families and families with strong local ties as reflected in marriage and residence patterns. Growing up in a collectivist household reduced the likelihood of subsequent emigration. The results are not explained by other relevant circumstances such socioeconomic status.

Due to selection, mass emigration generated a reduction in the stock of individualistic 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 over the course of the Age of Mass Migration but also up until today. The cultural change that took place during the Age of Mass Migration was thus sufficiently 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 industrialization, urbanization, and democratization in Scandinavia. Individualism was generally on the rise, in part due to these developments, but it seems conceivable that the collectivistic 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 contrasting the Scandinavian welfare model with American liberal capitalism.[33] Future research may identify the impact of culture on these developments.

This study may form the basis for future research in several directions. With the construction of valid historical cultural indicators at the individual and society level, cultural change and persistence can be investigated over long periods of time and underlying mechanisms identified in detail. For example, Knudsen (2018) studies the relevance of different channels of cultural transmission of collectivism at the individual child level across more than four million historical European families. Other research may study the impact of cultural traits and values on individual behavior and aggregate outcomes. Thus, further light may be shed on how culture at the same time evolves in and shapes social and economic structures of society.

[33]Today the Scandinavian countries are among the very top on the World Bank "Ease of Doing Business" ranking. The countries are characterized by low levels of corruption and the protection of human rights. At the same time, tax rates are among the highest in the world, and they are used to finance an extensive scheme of social insurrance policies. Labor unions are particularly strong and collective bargaining of wages is the norm. A famous characterization of Scandinavian culture is the "Law of Jante" which teaches people not to think too much of themselves. See discussions in Acemoglu et al. (2012) and Krugman (2015).

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] Acemoglu, D., Robinson, J. A., and Verdier, T. 2012. Can't We All Be More Like Scandinavians? Asymmetric growth and institutions in an interdependent world (No. w18441). *National Bureau of Economic Research.*

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

[6] Alesina, A., & Giuliano, P. 2015. Culture and institutions. *Journal of Economic Literature*, 53(4), 898-944.

[7] Andersson, D., Karadja, M., and Prawitz, E. 2017. Mass migration, cheap labor, and innovation. *Working Paper.*

[8] Arellano, M., and Bond, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), 277-297.

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

[10] Barry, H., Child, I. L., and Bacon, M.K. 1959. Relation of child training to subsistence economy. *American Anthropologist*, **61**(1), pp.51-63.

[11] Bazzi, S., Fiszbein, M. and Gebresilasse, M. M. 2018. Frontier Culture: The Roots and Persistence of Rugged Individualism in the United States. *Working Paper.*

[12] 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 (2019): 1764-1804.

[13] Bender, H. 2007. Kilderne på nettet til udvandringen fra Danmark. *Working paper.*

[14] Bentzen, J. S. (2019). Acts of God? religiosity and natural disasters across subnational world districts. *Economic Journal.*, 29.622 (2019): 2295-2321.

[15] Bhugra, D., & Becker, M. A. 2005. Migration, cultural bereavement and cultural identity. *World Psychiatry*, 4(1), 18-24.

[16] Bisin, A., and Verdier, T. 1998. On the cultural transmission of preferences for social status. *Journal of Public Economics*, **70**(1), 75-97.

[17] Bisin, A., and Verdier, T. 2000. "Beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. T*he Quarterly Journal of Economics*, **115**(3), 955-988.

[18] Bisin, A., and Verdier, T. 2001. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, **97**(2), 298-319.

[19] Bloothooft, G. and Groot, L. 2008. Name clustering on the basis of parental preferences. *Names,* **56**(1):111–163

[20] Bozeman, B. 2007. *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press.

[21] Buggle, J. C. 2017. Growing Collectivism: Irrigation, Group Conformity and Technological Divergence. *Working Paper.*

[22] Cavalli-Sforza, L. L., and Feldman, M. W. 1981. *Cultural transmission and evolution: a quantitative approach*. Princeton University Press.

[23] Cavalli-Sforza L. L., Menozzi, P., and Piazza, A. 1994. *The history and geography of human genes*, Princeton University Press

[24] Chen, C. C., Chen, X. P., & Meindl, J. R. 1998. How can cooperation be fostered? The cultural effects of individualism-collectivism. *Academy of management review*, 23(2), 285-304.

[25] Chowdhury, S., Sutter, M., and Zimmermann, K. 2018. Evaluating intergenerational persistence of economic preferences: A large scale experiment with families in Bangladesh. *Working Paper.*

[26] Darwin, C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* John Murray, London

[27] 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.

[28] Docquier, F., and Rapoport, H. 2012. Globalization, brain drain, and development. *Journal of Economic Literature*, **50**(3), 681-730.

[29] Dohmen, T., Falk, A., Huffman, D., and Sunde, U. 2011. The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, **79**(2), 645-677.

[30] Eisenbruch, Maurice. 1991. "From post-traumatic stress disorder to cultural bereavement: Diagnosis of Southeast Asian refugees." *Social Science and Medicine* (1991): 673-680.

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

[32] van Everdingen, Y. M., & Waarts, E. 2003. The effect of national culture on the adoption of innovations. *Marketing Letters*, 14(3), 217-232

[33] Fouka, V. 2019. Backlash: The unintended effects of language prohibition in US schools after World War I. *The Review of Economic Studies.*, 05

[34] Fryer, R.G. and Levitt, S.D., 2004. The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, **119**(3), pp.767-805.

[35] Galor, O., and Ömer Ö. 2016) The Agricultural Origins of Time Preference. *American Economic Review*, **106**(10): 3064-3103.

[36] Giuliano, P. and N. Nunn. 2019. Understanding Cultural Persistence and Change. *Working Paper.*

[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] Gorodnichenko, Y., and Roland, G. 2017. Culture, institutions, and the wealth of nations. *Review of Economics and Statistics*, **99**(3), 402-416.

[40] 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.

[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., Heine, S. J., & Norenzayan, A. 2010. The weirdest people in the world?. *Behavioral and brain sciences*, 33(2-3), 61-83.

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

[45] Indseth, T. 2006. Forestillinger om Amerika i Norge. *Masteroppgave, University of Oslo*

[46] 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.

[47] Kashima, E. S., and Kashima, Y., 1998. Culture and language the case of cultural dimensionsand personal pronoun use. *Journal of Cross-Cultural Psychology* **29** (3), 461–486.

[48] 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), pp.000-000.

[49] 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.

[50] Knudsen, A. S. B. 2018. The Intergenerational Transmission of Collectivistic Traits: Evidence Across Millions of Historical European Families. *Working paper.*

[51] Krugman, P. 2015, October 19. Something Not Rotten. *The New York Times*, pp. A23

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

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

[54] Long, J., and Ferrie, J. (2013). Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review*, **103**(4), 1109-37.

[55] Luterbacher, J., Daniel Dietrich, E. X., M. Grosjean, and H. Wanner. (2004). European Seasonal and Annual Temperature Variability, Trends, and Extremes Since 1500. *Science*, **303**(5663), 1499–1503.

[56] Macfarlane, A. 1978. *The origins of English individualism: the family property and social transition*. Wiley-Blackwell.

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

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

[59] Mill, R. 2012. Assessing Individual-Level Record Linkage Between Historical Datasets, *Stanford University Mimeo*.

[60] Minnesota Population Center. 2017. *North Atlantic Population Project: Complete Count Microdata*. Version 2.3 [dataset]. Minneapolis: Minnesota Population Center.

[61] Mokyr, J. 1992. Technological Inertia in Economic History. *The Journal of Economic History*, **52**(2): 325-338

[62] Morling, B. and Lamoreaux, M. 2008. Measuring culture outside the head: A meta-analysis of individualism–collectivism in cultural products. *Personality and Social Psychology Review*. **12**(3), 199-221.

[63] Morris, M. H., Avila, R. A., & Allen, J. 1993. Individualism and the modern corporation: Implications for innovation and entrepreneurship. *Journal of management*, 19(3), 595-612.

[64] Murrell, P. and Schmidt, M. 2011. The Coevolution of Culture and Institutions in Seventeenth Century England. *Working paper*.

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

[66] Semmingsen, I., and Haugen, E. I. 1978. *Norway to America: A History of the Migration*. U of Minnesota Press.

[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] Ogihara, Y., Fujita, H., Tominaga, H., Ishigaki, S., Kashimoto, T., Takahashi, A., Toyohara, K. and Uchida, Y. 2015 Are common names becoming less common? The rise in uniqueness and individualism in Japan. *Front. Psychol.* **6**:1490.

[69] Oishi, S. 2010. The psychology of residential mobility: Implications for the self, social relationships, and well-being. *Perspectives on Psychological Science*, **5**(1), 5-21.

[70] Oishi, S., Lun, J., and Sherman, G. D. 2007. Residential mobility, self-concept, and positive affect in social interactions. *Journal of Personality and Social Psychology*, **93**(1), 131.

[71] Olsson, O. and Paik, C. 2016. Long-run cultural divergence: Evidence from the Neolithic Revolution, *Journal of Development Economics*. **122**,(C), 197-213

[72] 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.

[73] Schwartz, S. H. 2004. Mapping and interpreting cultural differences around the world. *International Studies in Sociology and Social Anthropology*, 43-73.

[74] Schwartz, S. H. 1994. *Beyond individualism/collectivism: New cultural dimensions of values*. Sage Publications, Inc.

[75] Sønnichsen, O. 2013. *Rejsen til Amerika*. Gyldendal.

[76] Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X. and Kitayama, S.

[77] (2014). Large-Scale Psychological Differences within China Explained by Rice versus Wheat Agriculture. *Science*, **344** (6184), 603–608.

[78] Tan, P. K. 2004. Evolving naming patterns: anthroponymics within a theory of the dynamics of non-Anglo Englishes. *World Englishes*, **23**(3), 367-384.

[79] Taylor, A. M., and Williamson, J. G. 1997. Convergence in the age of mass migration. *European Review of Economic History*, **1**(1), 27-63.

[80] Taylor, M. Z., & Wilson, S. 2012. Does culture still matter?: The effects of individualism on national innovation rates. *Journal of Business Venturing*, 27(2), 234-247

[81] Turner, F. J. 1920. *The Frontier in American History* (New York, 1920), The Significance of Sections in American History

[82] 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.

[83] Uz, I. 2015. The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology*, **46** (3), 319—335

[84] Van de Vliert, E. 2011. Climato-economic origins of variation in ingroup favoritism. *Journal of Cross-Cultural Psychology*, **42** (3), 494–515.

[85] 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.

[86] 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.

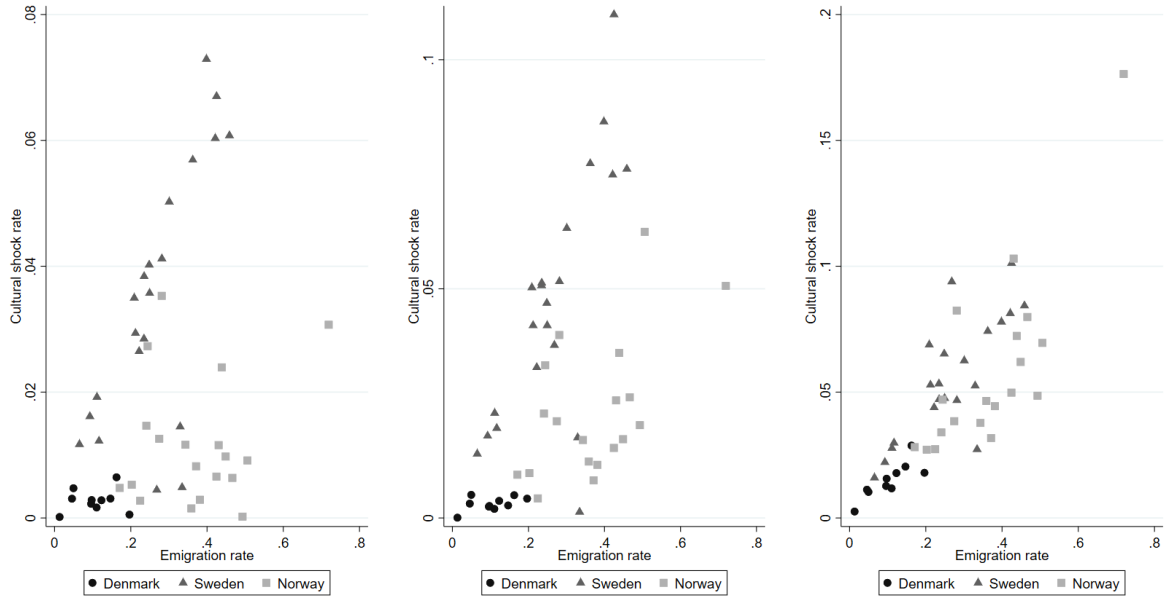
[87] Wagner III, J. A. 1995. Studies of individualism-collectivism: Effects on cooperation in groups. *Academy of Management journal*, 38(1), 152-173.

[88] Welzel, C. 2013. *Freedom Rising: Human Empowerment and the Quest for Emancipation*. New York: Cambridge University Press.

[89] Zweigenhaft, R. L. 1981. Unusual names and uniqueness. *The Journal of Social Psychology*, **114**(2), 297-298.

A Appendix: Additional Figures and Tables

Figure A.1. Rate and Cultural Shock of Emigration, 1850-1920



Notes: Scatterplot of the total rate of emigration and the relative drop in district first name commonness associated with selective emigration. Measures represent alternatives to the baseline (based on the actual spelling of first names, all emigrants and stayers, and the names distribution in the census closest to birth year). Instead, the illustrated measures are based on the phonetic spelling of first names (first plot), the names distribution in the census closest to emigration year (second plot), and emigrants and stayers that carry a name found in both census and passenger lists (third plot).

Table A.1. Validation using contemporary indicators: Countries and sources

		Country level		District level		
Continent	Country	Top one	Top ten	Top one	Top ten	Source
Africa	South Africa	x	x			Imbizo Centre
Americas	Argentina	x				Registro Civil de la Ciudad de Buenos Aires (via lanacion.com) (Only Buenos Aires)
Americas	Canada	x	x	x	x	Governments of Alberta, British Columbia, Manitoba, Nova Scotia, Ontario, Québec, Saskatchewan, and Northern Territories
Americas	Chile	x	x			Chile Registro Civil
Americas	Colombia	x	x			Registraduria Nacional del Estado Civil
Americas	Peru	x				Registro Nacional de Identificación y Estado Civil (via elcomercio.pe) (Only 3 first months of 2015)
Americas	United States	x	x	x	x	US Social Security
Asia	Israel	x	x			Central Bureau of Statistics Israel (Only Jewish population)
Asia	Philippines	x	x			Philippine Statistics Authority
Europe	Austria	x	x	x	x	Statistik Austria
Europe	Belgium	x	x	x	x	Statbel (Algemene Directie Statistiek-Statistics Belgium)
Europe	Bulgaria	x				National Statistical Institute (NSI) (via nancy.cc)
Europe	Czech Republic	x	x	x	x	Ministry of the Interior of the Czech Republic
Europe	Denmark	x	x	x	x	Danmarks Statistik
Europe	Estonia	x	x			Estonia Ministry of the Interior (via nancy.cc)
Europe	Finland	x	x			Population Register Centre
Europe	France	x	x	x	x	Institut national de la statistique et des études économiques (INSEE)
Europe	Germany	x	x			blog.beliebte-vornamen.de (independent collection of city data)
Europe	Hungary	x	x			Belügyminisztérium Nyilvántartások Vezetéséért Felelos Helyettes Államtitkárság (via behindthename.com)
Europe	Iceland	x	x			Statistics Iceland
Europe	Ireland	x	x	x	x	The Central Statistics Office of Ireland (SCO)
Europe	Italy	x	x	x	x	Istat - Istituto Nazionale di Statistica
Europe	Latvia	x	x			Office of Citizenship and Migration Affairs (via bnn-news.com)
Europe	Lithuania	x	x			Population Register Service (via vardai.vlkk.lt)
Europe	Malta	x	x			Malta National Statistics Office (via nancy.cc) (Phonetic spellings combined)
Europe	Moldova	x	x			Moldova Civil Status Service (via noi.md)
Europe	Netherlands	x	x	x	x	Meertens Institut
Europe	Norway	x	x	x	x	StatBank Norway
Europe	Poland	x	x			Ministry of Interior and Administration (via behindthename.com)
Europe	Portugal	x	x			Instituto dos Registos e do Notariado (via behindthename.com)
Europe	Romania	x	x			Ministry of Administration and Interior (via behindthename.com)
Europe	Slovakia	x	x			Interior Ministry (via spectator.sme.sk)
Europe	Slovenia	x	x			Republic of Slovenia Statistical Office RS
Europe	Spain	x	x	x	x	Instituto National de Estadística
Europe	Sweden	x	x	x		Statistics Sweden
Europe	Switzerland	x	x	x	x	Federal Statistical Office Switzerland
Europe	United Kingdom	x	x	x		Office for National Statistics
Oceania	Australia	x	x	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 along with secondary source websites in parentheses where relevant.

Table A.2. Cross-country validation using contemporary indicators

	(1)	(2)	(3)	(4)
Dep. Variable	Hofstede Individualism Index			
Baby name commonness	-0.465***	-0.444***	-0.466***	-0.367***
	(0.093)	(0.098)	(0.074)	(0.095)
Uncommonness measure	Baseline	Baseline	Baseline	Baseline
Fixed Effects	None	Continent	Continent	Continent
Controls	N	N	Y	Y
Add. Hofstede Controls	N	N	N	All
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 baby first name commonness, defined as the share of children born in 2015 that were given one of the 10 most popular boys or girls names in their country in this year. Controls include the log number of newborns, ethnic fractionalization, genetic diversity, and a latin language dummy. Additional Hofstede cultural variables include Power Distance, Masculinity, Uncertainty Avoidance, Long Term Orientation, and Indulgence. All available data across the globe is included in these regressions. 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 A.3. Cross-country validation with alternative indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Survey Based Measures				Linguistic Characteristics		
Dep. Variable	Embeddedness	Affective Autonomy	Intellectual Autonomy	In-Group Favouritism	Social Tightness	Pronoun Drop (KK 1998)	Pronoun Drop (AD 2014)	Subject Prominence
Baby name uncommonness	0.550**	-0.728***	-0.342	0.496***	0.467**	0.501***	0.666**	-0.378**
	(0.215)	(0.223)	(0.226)	(0.148)	(0.172)	(0.131)	(0.299)	(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. Baby name commonness is calculated as the share of newborns recieving one of 10 most popular male or female names of their birth year in 2015. 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). Controls include continent fixed effects, log number of newborns, ethnic fractionalization, genetic diversity, and a latin language dummy. All variables have been standardized before regression. Robust standard errors are shown in the parentheses with the following significance levels:*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4. Within-country validation with contemporary indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(7)	(8)
	Emancipative Values		Personal Autonomy		Gender Equality		Lifestyle Liberty		Voice of the People	
Baby name commonness	-0.052**	-0.398***	-0.043**	-0.271**	-0.019	-0.263***	-0.048**	-0.349***	0.006	-0.055*
	(0.024)	(0.112)	(0.020)	(0.111)	(0.012)	(0.073)	(0.023)	(0.117)	(0.006)	(0.030)
Countries	13	3	13	3	13	3	13	3	13	3
Observations	52,822	6,301	56,954	7,503	17,719	2,199	54,290	6,347	59,308	7,670
R-squared	0.24	0.37	0.11	0.12	0.15	0.07	0.19	0.27	0.05	0.02

Notes: The unit of observation is an individual surveyed in WVS/EVS. "Emancipative Values" is calculated according to Welzel (2013) at the individual level based on questions regarding personal autonomy, gender equality, lifestyle liberty, and freedom of voice. These variables represent the dependent variables in this table. Odd-numbered columns include all countries available in the sample, whereas even-numbered columns restrict the analysis to Norway, Denmark and Sweden. Controls include country fixed effects, the log number of newborns, respondent age, age squared and gender. All variables have been standardized before regression. Robust standard errors, clustered at the district level, are shown in the parentheses. Significance levels are: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5. Validation using Swedish newspaper language, 1780-1900

Dep. Variable	(1)	(2)	(3)	(4)
	Relative Singular Pronoun Use			
	All pronouns		Pronoun + verb	
First name cmn.	-9.682**	-2.236**	-10.383**	-2.514**
	(3.561)	(1.009)	(3.876)	(1.084)
Commonness measure	Top 10	Log share	Top 10	Log share
Observations	70	70	70	70
R-squared	0.93	0.93	0.94	0.94

Notes: The unit of observation is a decade and newspaper. Relative singular pronoun use equals the log ratio between singular and plural personal pronouns used in 20 Swedish local newspapers. The measure considers all pronouns in columns (1) and (2), and pronouns that are associated with a verb in columns (3) and (4). This is regressed on the average first name commonness of people born in the given decade and region of the newspaper. In columns (1) and (3) this is given as the share of the birth cohort with names among the 10 most popular in the cohort. Otherwise it is the log average share of a cohort that share a given first name. All regressions control for the amount of newspaper text (log word count), population size, and birth cohort size. Newspaper fixed effects are included as well. Robust standard errors are reported in parentheses. Significance levels are: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6. Most popular children names in select censuses

	Female		Male		Female		Male	
Rank	Name	Name share	Name	Name share	Name	Name share	Name	Name share
Denmark		Year 1845				Year 1880		
1	Ane	15%	Jens	11%	Ane	12%	Jens	10%
2	Maren	9%	Niels	10%	Karen	5%	Hans	9%
3	Karen	8%	Hans	10%	Anna	5%	Niels	9%
Norway		Year 1865				Year 1910		
1	Anne	6%	Ole	9%	Anna	4%	Ole	3%
2	Anna	4%	Hans	5%	Astrid	3%	Johan	3%
3	Karen	3%	Johan	5%	Gudrun	2%	Karl	2%
Sweden		Year 1880				Year 1900		
1	Anna	16%	Johan	12%	Anna	12%	Karl	13%
2	Maria	5%	Carl	11%	Ester	4%	Johan	8%
3	Emma	4%	Karl	9%	Elsa	3%	Gustaf	6%

Notes: Country level measures based on children below the age of 10 in select censuses.

Table A.7. Cross-validation of censuses and passenger lists

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Change in Population Census					
Emigrants	-2.583***	-2.575***	-2.621***	-0.556***	-0.553***	-0.574***
	(0.093)	(0.093)	(0.094)	(0.115)	(0.115)	(0.116)
Total Number in Census				-0.162***	-0.162***	-0.162***
				(0.007)	(0.007)	(0.007)
Fixed Effects	None	Census	Cohort, District, Census	None	Census	Cohort, District, Census
Observations	848,907	848,907	848,907	848,907	848,907	848,907
R-squared	0.23	0.24	0.25	0.43	0.43	0.44

Notes: OLS regressions comparing the change from one population census to the next in the number of people born in the same cohort, residing in the same district and carrying the same first name with the number of emigrants with the same characteristics. Robust standard error in parentheses with significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.8. Cross-validation of emigration data (Norway and Sweden)

	(1)	(2)	(3)	(4)
Dep. Var.	Emigrants (national accounts)			
Passenger list emigrants	1.392***	1.108***		
	(0.074)	(0.073)		
Passenger list emigrants			1.555***	1.242***
(with personal information)			(0.082)	(0.086)
Fixed Effects	Country	District	Country	District
Observations	240	240	240	240
R-squared	0.73	0.90	0.72	0.90

Notes: OLS regressions comparing the absolute number of emigrants recorded each decade 1860-1910 across 40 Swedish and Norwegian subnational districts. Robust standard errors in parentheses with significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.9. Cross-validation of emigration data (Denmark)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Emigrants (Hamborg lists)				Emigrants (joint lists)	
Passenger list emigrants	0.023	-0.049			1.001***	0.957***
	(0.022)	(0.040)			(0.020)	(0.049)
Joint lists emigrants			0.051***	-0.015		
			(0.018)	(0.037)		
Fixed Effects	Country	District	Country	District	Country	District
Observations	60	60	60	60	60	60
R-squared	0.01	0.46	0.07	0.42	0.97	0.97

Notes: OLS regressions comparing the absolute number of Danish emigrants recorded in different passenger lists for each decade 1860-1910. Robust standard error in parenthesis with significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.10. Selective emigration: Alternative commonness measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. var.	Emigration								
First name cmn	-0.022***	-0.007***	-0.008***	-0.007***	-0.008***	-0.009***	-0.009***	-0.008***	-0.009***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Commonness measure	Top 10 popularity	Using full first name	Using phonetic spelling	Using the first letter	Using the first 3 letters	Among officially recognized names	Among non-rare names	Country level cmn.	Municipality level cmn.
Controls	X	X	X	X	X	X	X	X	X
Observations	936,125	936,125	936,125	936,125	936,125	871,354	922,806	936,125	935,850
R-squared	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17

Notes: The sample consists of males below the age of 15 in the earliest censuses (DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, SE-1900). The emigration indicator equals one if the individual is observed in the passenger lists and zero if observed the later censuses (DK-1901, NO-1910, SE-1910). Controls include age in census, number of siblings, birth order fixed effects, census X birth decade X birth district fixed effects, last name commonness, additional first name characteristics (string length, number of names, rare names, officially recognized names, and name shared with a family member), and household controls (urban, farming, employing servants, family size, and for the mother and father: being on poor relief, owning assets, being in the labor force, having skilled work, having non-manual work, and being foreign). Standard errors clustered at the district level. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.11. Selective emigration: Additional fixed effects

	(1)	(2)	(3)	(4)
Dep. var.	Emigration			
First name commonn. (ln)	-0.009***	-0.009***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Additional fixed effects	Birth municipality	Last name	Father's occupation	Fathers' wealth
Controls	X	X	X	X
Observations	935,836	917,838	936,125	936,125
R-squared	0.18	0.24	0.18	0.18

Notes: The sample consists of males below the age of 15 in the earliest censuses (DK-1845, DK-1880, NO-1865, NO-1900, SE-1880, SE-1900). The emigration indicator equals one if the individual is observed in the passenger lists and zero if observed the later censuses (DK-1901, NO-1910, SE-1910). Controls include age in census, number of siblings, birth order fixed effects, census X birth decade X birth district fixed effects, last name commonness, additional first name characteristics (string length, number of names, rare names, officially recognized names, and name shared with a family member), and household controls (urban, farming, employing servants, family size, and for the mother and father: being on poor relief, owning assets, being in the labor force, having skilled work, having non-manual work, and being foreign). Standard errors clustered at the district level. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.12. Selective emigration: Group-level results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. variable	Rate of emigration							
Log first name cmn.	-0.002***	-0.002***	-0.004***	-0.002***	-0.002***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Sample	Full			Children in census				
Individual controls	X	X	X	X	X	X	X	X
Add. first name char.			X			X	X	X
Sibling controls							X	X
Household char.								X
Census X birth decade	X			X				
Census X b. decade X district		X	X		X	X	X	X
Mean(s.d.) dep. var.	.05(.08)	.05(.08)	.05(.08)	.04(.07)	.04(.07)	.04(.07)	.04(.07)	.04(.07)
Mean(s.d.) indep. var.	-4.1(1.8)	-4.1(1.8)	-4.1(1.8)	-4.3(1.8)	-4.3(1.8)	-4.3(1.8)	-4.3(1.8)	-4.3(1.8)
Observations	642,299	642,299	642,299	304,650	304,650	304,650	304,650	304,650
R-squared	0.21	0.38	0.40	0.16	0.29	0.31	0.31	0.31

Notes: Results from group-level estimations, where the rate of emigration among people that share the same birth decade, gender, residence district and first name is regressed on the commonness of that name. In columns (1)-(3) all emigrants are considered and in columns (4)-(8) only those that were below the age of 15 in the census are considered. First name commonness is calculated using the census in which this cohort was youngest, and regressions are weighted according to the size of these groups. Control variables represent averages and include individual (age and gender), additional first name characteristics (length, number of first names, rare names, officially recognized names, and shared by another family member), sibling controls (number and birth order), and household controls (urban location, being engaged in farming, having servants employed, family size, and separately for the mother and father: being on poor relief, owning assets, being in the labor force, having skilled work, having non-manual work, and being foreign). Standard errors clustered at the district level. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.13. Assimilation of Scandinavian immigrants in the US

Dep. var.	(1) No English spoken	(2) No English spoken	(3) Spouse same nationality	(4) Spouse same nationality	(5) State concentration of own nationality	(6) State concentration of own nationality	(7) First name ethnic distinctiveness of child born in the US	(8) First name ethnic distinctiveness of child born in the US	(9) First name ethnic distinctiveness of child born in the US	(10) First name ethnic distinctiveness of child born in the US
Log first name comn.	0.0016***	0.0011***	0.0039***	0.0035***	0.0003***	0.0002***	0.0014***	0.0011***	0.0009***	0.0006**
	0	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years in the US	-0.0079***	-0.0082***	-0.0041***	-0.0040***	0.0001***	0.0000	-0.0015***	-0.0016***	-0.0023***	-0.0024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	All	All	All	All	Fathers	Fathers	Mothers	Mothers
Additional controls	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	47,758	47,758	47,758	47,758	47,758	47,758	37,335	37,335	32,371	32,371
R-squared	0.18	0.19	0.54	0.54	0.15	0.21	0.05	0.05	0.06	0.06
Mean of dep. var.	0.0773	0.0773	0.458	0.458	0.0268	0.0268	0.0561	0.0561	0.0682	0.0682
St.dev. of dep. var.	0.267	0.267	0.498	0.498	0.0245	0.0245	0.128	0.128	0.145	0.145

Notes: OLS regressions. Unit of observation is a Scandinavian immigrant in the Unites States 1900 and 1910 population census. Log first name commonness (mean -6.6, s.d. 3.0) is measured as the share of a birth decade cohort in the 1880 Danish and the 1900 Swedish and Norwegian censuses that carried the same first name. Non-assimilation is given by the lack of English proficiency, by marrying someone of own nationality, settling in a state with a high concentration of own nationality, and by giving a child a more distinctively Scandinavian sounding name (compared to the entire US census population of 1900 and 1910). All regressions contain controls for age, age squared, a gender, marriage, marriage in the US, census as well as cohort fixed effects. In the last four columns the number of siblings, age, age squared and gender of the child is also controlled for. Additional controls include first name length, indicators for living on a farm or in an urban area, the Duncan Socioeconomic index, and the IPUMS occupational income score. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.14. Cultural change: Components of the individualism index

Dep. variable	(1) Personal autonomy	(2) Personal autonomy	(3) Lifestyle liberty	(4) Lifestyle liberty	(5) Gender equality	(6) Gender equality	(7) Voice of people	(8) Voice of people
Historical first name cmn.	-0.046	-0.015	-0.126*	-0.078	-0.000	-0.057	-0.043*	-0.040
	(0.031)	(0.043)	(0.065)	(0.068)	(0.041)	(0.056)	(0.023)	(0.024)
Cumulative shock	-0.677**	-0.797**	-0.924***	-1.427***	-0.455	-0.218	-0.408**	-0.244
	(0.264)	(0.369)	(0.340)	(0.462)	(0.524)	(0.727)	(0.180)	(0.174)
Additional controls		X		X		X		X
Observations	7,503	7,503	6,333	6,333	2,199	2,199	7,676	7,676
R-squared	0.12	0.12	0.28	0.28	0.07	0.07	0.10	0.10

Notes: The unit of observation is an individual surveyed in the World or European Values Surveys. The dependent variables are sub-components of the Welzel (2013) "Emancipative Values" indicator of individualism. Historical first name commonness is the average popularity of first names given to children (< 10 years old) in the earliest census for each country (DK-1845, NO-1865, SE-1880), calculated at the district level (for 50 districts). The cumulative shock equals the drop in first name commonness from selective emigration (see main text for explanation). All regressions include country, historical census year, and WVS/EVS survey year fixed effects. Individual controls are age, age squared, and gender. Geographical controls are log area, soil suitability for agriculture, share of arable land, and (minus log) distance from the coast. Historical controls are average last name commonness, family size, children per mother, and the propensity to live in urban areas, have servants employed, hold assets, hold skilled work, hold non-manual work, and be named after a family member. The shock of emigration to last name commonness and gender balance is included too. Standard errors clustered at the district level are shown in the parentheses with the following significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.15. Cultural change: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A	First name commonness 1901/1910							
Historical first name cmn.	0.729***	0.714***	0.684***	0.729***	0.675***	0.666***	0.664***	0.721***
	(0.055)	(0.100)	(0.084)	(0.096)	(0.107)	(0.100)	(0.101)	(0.106)
Cumulative shock	2.611**	1.256	2.031**	0.482	2.839*	2.586**	0.391	1.555
	(1.178)	(1.517)	(0.915)	(1.510)	(1.641)	(1.205)	(1.197)	(1.011)
Observations	121	50	50	50	50	50	50	50
R-squared	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97
Panel B	First name commonness 2015							
Historical first name cmn.	0.272***	0.293***	0.380***	0.352***	0.282**	0.281**	0.135	0.130
	(0.052)	(0.096)	(0.093)	(0.101)	(0.118)	(0.106)	(0.106)	(0.109)
Cumulative shock	2.610**	2.867**	-0.866	0.121	2.962*	2.492*	3.314**	3.227***
	(1.042)	(1.284)	(0.864)	(1.269)	(1.670)	(1.335)	(1.263)	(1.084)
Observations	171	50	50	50	50	50	50	50
R-squared	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Panel C	Individualism index (WVS/EVS)							
Historical first name cmn.	-0.073***	-0.058*	-0.063*	-0.057*	-0.058	-0.052	-0.048	-0.001
	(0.016)	(0.034)	(0.032)	(0.031)	(0.040)	(0.033)	(0.031)	(0.027)
Cumulative shock	-0.163	-0.619**	-0.396**	-0.719***	-0.454	-0.594***	-0.824***	-0.643***
	(0.297)	(0.270)	(0.194)	(0.232)	(0.389)	(0.212)	(0.267)	(0.233)
Observations	22,002	6,303	6,303	6,303	6,303	6,303	6,303	6,303
R-squared	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.35
Censuses	All	Earliest	Earliest	Earliest	Earliest	Earliest	Earliest	Earliest
Shock measure	Baseline	Phonetic	Alt. census link	Intensive margin	Danish emigration 1	Danish emigration 2	Historical controls	Geo. controls

Notes: Cross-district OLS regressions, predicting cultural differences towards the end of the Age of Mass Migration, panel A and B, and today, Panel C. Historical first name commonness is the average popularity of first names given to children (< 10 years old) in the earliest census for each country (DK-1845, NO-1865, SE-1880), calculated at the district level. The cumulative shock equals the increase in first name commonness directly associated with selective emigration (see main text for explanation). All regressions control for country fixed effects, initial population levels, and basic characteristics of the chidren whose first names are used to proxy initial collectivist traits (gender ratio, average number of first names, and first name string length). Additional controls include the total rate of emigration over the period, the commonness of last names among the children in the early censuses, and the shock of emigration to last name commonness. In panel A the dependent variable is average first name commmonness among children in the 1901/1910 censuses, and in panel B it is the commonness of baby names in 2015 In panel C the dependent variable is the Welzel (2013) measure of "Emancipative Values" in the WVS and EVS. The regressions here are estimated at the individual respondent level, including individual characteristics (age, age squared, gender), survey year fixed effects, and standard errors clustered at the district level. Standard errors are shown in the parentheses with the following significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

B Appendix: Linking Migrants with Census Records

I link male emigrants and non-emigrants to the population census in which they were below the age of 15. Emigrants are identified from passenger lists and non-emigrants as those still living in the country according to later censuses. I match emigrants and remaining out-migrants to their childhood household using the following procedure:

1. A sample of potential links is constructed based on the phonetic spelling (truncated to the first four letters) of first and each last name, a two-year band around the birth year, and district of residence (emigrants) or birth (non-migrants). These are binding matching criteria.

2. Migrants that share the exact characteristics with one or more individuals in the census are identified as either exact matches or failed matches respectively. In terms of district of residence/birth similarity this step concerns the village (within district) level.

3. From here follows an iterative process where each matching criteria (similarity in first name, last name, birth year, and district of residence or birth) is weakened. The birth year is allowed to differ with one or two years. Non-perfect string similarity between first and last names is allowed (with Jaro-Winkler measures of 0.95, 0.9, 0.85., 0.8, and 0.75). Matching district of residence/birth is allowed to be at a higher level than the village: Sub-district and district.