THE UNIVERSITY OF CHICAGO

FORGING APARTHEID: MARKET ACCESS AND THE RISE OF NATIONAL IDENTITY IN COLONIAL SOUTH AFRICA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE IRVING B. HARRIS

GRADUATE SCHOOL OF PUBLIC POLICY STUDIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

LAURA MONTENEGRO HELFER

CHICAGO, ILLINOIS

MARCH 2024

Contents

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

2 Historical Overview **11**

3 Conceptual Framework **13**

3.a Market Integration, Social Identity Scale, and the Role of Culture 14

3.b British or Afrikaner? Choosing a National Identity amid Racial Mobility . . 16

4 Data **19**

4.a Market Access 21

4.a.1 Defining Transportation Cost Parameters 21

4.a.2 Transportation GIS Network 23

4.a.3 Computing Market Access 26

4.b National Identity Variables 28

4.b.1 Children's First Name Etymologies 29

4.b.2 Newspaper Language 34

4.b.3 Grassroots Organization Meetings (Cape of Good Hope Colony) . . . 35

4.b.4 House of Assembly Members (Cape of Good Hope Colony) 37

4.c Mechanisms: Wages, Agricultural Production, and Land Grants 39

5 Empirical Strategy **42**

5.a Discussion on Further Potential Identification Challenges (and Some Solutions for the Near Future) 43

6 Results **46**

6.a Estimated Impact of Market Access on National Identity 46

6.a.1 Estimated Impact of Market Access on Children's First Names 46

6.a.2 Estimated Impact of Market Access on Newspaper Language 47

6.a.3 Estimated Impact of Market Access on Grassroots Organizations (Cape) 48

6.a.4 Estimated Impact of Market Access on MPs Last Names (Cape) . . . 49

6.a.5 Heterogeneities by Initial Populations: White Protestant, White DRC, and Non-White Populations 52

6.b Addressing Identification Challenges 52

6.b.1 Controlling for Infrastructure Placement 52

6.b.2 Placebo Infrastructure 56

7 Mechanisms **58**

7.a Estimated Impact of Market Access to Different Population Profiles: Race and Belief Systems (Religion) 59

7.a.1 Estimated Impact of Market Access to Different Racial Profiles . . . 60

7.a.2 Estimated Impact of Market Access to Belief Systems and Racial Profiles 64

7.b Changes in Relative Status 67

7.b.1 Estimated Impact of Market Access on Racial Wage Gaps 68

7.b.2 Estimated Impact of Market Access on Traditional Agricultural Production by Racial Profile 69

8 Concluding Remarks **73**

A White Settler's Maximization Problem **74**

B Additional Figures and Tables **77**

C Selected Primary Sources on the Cape of Good Hope Colony **92**

C.1 National Archives of South Africa: Cape Town Archive Repository (KAB) . 92

C.1.1 Census District Maps (1904). Compiled from maps filed in the Surveyor General's Office 92

C.1.2 Blue Books and Statistical Registers (1869-1909). British Foreign & Commonwealth Office: 95

C.2 Stellenbosch University 95

C.2.1 Railway Records 95

References **96**

List of Figures

1 Georeferenced Spatial Object of South Africa 20

2 Railroad Network Expansion (1869-1909) 24

3 iGraph and Dijkstra Path Calculations 27

4 Market Access Variable in 1869, 1909, and Change Between Years 28

5 MPs Last Name Etymologies 38

B1 Road Construction Heat Map (1869-1909) 77

B2 Roads Intersections & Expansion of Telegraphs 78

B3 Baseline Racial Wage Gaps (Cape of Good Hope) 79

List of Tables

1 GIS Network (1869-1909) 22

2 Afrikaans Naming Rules (Examples) 30

3 Ten Common Children's Names (1870-1910) 33

4 Summary Statistics of Outcome Variables 41

5 Estimated Impact of Market Access on National Identity Choice 51

6 Estimated Impact of Market Access on National Identity Choice Controlling for Local Infrastructure Placement 54

7 Estimated Impact of Market Access on National Identity Choice Controlling for Placebo Infrastructure 57

8 Estimated Impact of Market Access on National Identity Choice Separating by Population Racial Profiles 63

9 Estimated Impact of Market Access on National Identity Choice by Religion and Non-White Populations 66

10 Estimated Impact of Market Access on Racial Wage Gaps by Job Type . . . 70

11 Estimated Impact of Estimated Impact of Market Access on Traditional Agricultural Production 72

B1 Estimated Impact of Market Access on Grassroots Organizations Meetings' Topics 80

B2 Estimated Impact of Market Access (total) on MPs Occupations 81

B3 (*first*) Estimated Impact of Market Access on National Identity Choice Heterogeneous Effects by Initial White Populations' Religions and Non-White Populations 82

B4 Estimated Impact of Market Access on Baby Names (Interactions by Race) . 84

B5 Estimated Impact of Market Access on Newspaper Language (Interactions by Race) 85

B6 Estimated Impact of Market Access on Meetings (Interactions by Race) . . . 86

B7 Estimated Impact of Market Access on MPs Names (Interactions by Race) . 87

B8 Estimated Impact of Market Access on Baby Names (Interactions by Religion) 88

B9 Estimated Impact of Market Access on Newspaper Language (Interactions by Religion) 89

B10 Estimated Impact of Market Access on Meetings (Interactions by Religion) . 90

B11 Estimated Impact of Market Access on MPs Names (Interactions by Religion) 91

1 Introduction

Do markets make or break nations? Market integration has been central to the consolidation of nations (Anderson, 1983; Pomeranz, 2000; Beckert, 2015). By extending pro-social behaviors beyond immediate social networks (Henrich, 2020), deepening markets have increased the scale of social organizations, contentious politics, and even the state (Nagel & Olzak, 1982; Tilly, 1986; Olzak, 1992; Tarrow, 1994), allowing nations that coordinate people under strong *imagined communities* with shared goals to grow as well (Deutsch, 1953; Weber, 1976; Gellner, 1983; Hroch, 1985; Anderson, 1983; Hobsbawm, 1990). Most successful nations are those which people massively identify with, supporting preferred norms and institutions' at the state level (Akerlof & Kranton, 2000, 2002, 2010).

The effect of market integration on people's national identity choice is however theoretically ambiguous. While roads turned *peasants into Frenchmen* by integrating heterogeneous regions into one culturally cohesive nation (Weber, 1976; Blanc & Kubo, 2021), deepening markets broke the Ottoman Empire into several nations—coining the phenomenon *balkanization* (Hechter, 2000; Jenne, 2004). This puzzle continues to play out today as scholars stress how globalization fosters cultural convergence in one setting (Maystre et al., 2014) and bolsters nationalism and divergence in another (Shayo, 2009, 2020). The extent to which market integration homogenizes populations or activates *boundaries* between them is a question open to empirical testing (Barth, 1969).

This paper aims to explore this question by empirically assessing the impact of market integration on people's national identity choice in colonial South Africa. This multiracial

society constitutes a unique setting to analyze such dynamics. The rollout of a massive infrastructure network following a mineral revolution integrated near-autarkic white settler societies structured around strict racial hierarchies to British-led global markets, inducing white settlers to identify with large-scale nations that proposed opposing normative views on the rules that should govern racial relations.[1] The dominant British supported relative racial integration under market-oriented institutions while the nascent Afrikaner advocated for institutionalized racial "separatness", or *Apartheid*.[2]

In recent decades, the social science literature has attributed nations' success and ubiquity as the world's leading political units to their versatility (Hannan, 1979; Nielsen, 1985; Olzak, 1992; Wimmer, 2008; Munashinghe, 2018). These social groups strategically tailor norms and institutions to coordinate people's access or restriction to social categories and roles based on malleable cultural criteria—such as race, language, religion, or adherence to common myths—frequently at non-members' expense (Anderson, 1983).[3] To succeed at inducing mass identification, a nation must leverage the cultural criteria that most broadly aligns the interests of heterogeneous populations in the social anonymous network. If there are mass disalignments regarding such criteria, a national *boundary* may activate.

[1]Railroads were virtually non existent in 1870, amounted to 6'900 *km* in 1890, 12'000 *km* in 1910, and 19'000 *km* in 1930 (own calculations). While the USA and India had the most *km* of rail by 1912 (397,387 *km* and 53,919 *km* vs. 12,552 *km* in South Africa), the *km* of rail network per inhabitant in South Africa was close to that of the USA (2.10 vs. 2.67), both of them lower than Argentina (4.27) (Fourie & Herranz-Loncan, 2015).

[2]The British empire abolished the slave trade in 1807 (Slave Trade Act) and overall slavery in 1833 (Slavery Abolition Act). Starting in 1854 the Cape of Good Hope Colony enfranchised people based on property, income, and gender requirements, not race.

[3]These cultural characteristics overlap with ethnic characteristics. Nations and ethnic groups present similar characteristics, which is why they are sometimes analyzed as equivalent social groups. In some views, nations are ethnic groups that develop nationalism—the ideology that stresses that the boundary of the nation should be congruent with that of the state (Nagel, 1995; Gellner, 1983).

In this light, I first hypothesize that locations with higher increases in market integration induce white settlers to identify with large-scale national groups. The activation of a within-white cultural boundary—i.e. British and Afrikaner—may suggest that disagreements on cultural criteria exist. Second, I theorize that dissent over racial criteria drive this divide. Market integration's differential impacts on racial mobility may unevenly affect white settlers' social utility that regulates the normative acceptance of racial integration. If mobility decreases social utility, the marginal value of institutionalizing non-white populations' repression may increase, inducing Afrikaner identification. Otherwise, the marginal cost of deviating from the dominant British may rise, inducing British identification.

To estimate the causal effects of market integration on people's national identity choice I build a set of aggregate "market access" measures, a reduced-form expression derived from general equilibrium trade theory that captures each location's (district's or sub-district's) costs of trade (Donaldson & Hornbeck, 2016). I use a variety of historical data to trace the yearly evolution of road and railroad networks between 1869 and 1909 and calculate the lowest-cost location-to-location freight routes between markets of different sizes. I exploit geographic and time variation in local market access increases to estimate the impact on identity choice, using location and time fixed effects that flexibly control for time-variant general shocks and time-invariant local differences across locations.

I measure national identity using a range of primary sources such as church records, newspapers, meeting advertisements, and public background records, leveraging the fact that cultural boundaries typically involve the use of distinct existing or newly created languages

to separate nations' members (Anderson, 1983; Harari, 2015).[4] These sources reveal nations' overarching reach into people's private and public lives—the home and the world. First, I measure parents' choice of children's first names by developing an original linguistic function that captures the gradual standardization of the créole Afrikaans language tailored to orthographical rules typically observed in naming practices; I also measure first names in standardized English and other European languages[5].

Second, I identify the language of newspaper publications: the nascent Afrikaans, English, and other European, African, and Indian languages. This measure allows me to observe how different sections of civil society adopt either common or different national languages for communication in the public arena. Third, I track meetings organized by two distinct grassroots organizations: one that is an explicit proponent of Afrikaner nationalism and another that promotes adhering to British institutionality. Finally, I measure the last name etymologies of members of parliament, since contemporary accounts state that supporters of British integration voted for MPs with English last names while proponents of the Afrikaner nation favored MPs with Dutch last names.

Results show that aggregate market access increases induce white settlers to identify with either the British or the Afrikaner nation. A one standard deviation greater increase in local market access between 1869 and 1909 heightens parent's choice of children's first names in Afrikaans by 10% and in English by 5%, decreasing Dutch names by 6%. Similarly, South African newspaper publications in Afrikaans increase in 24% and in English by 23%

[4] A phenomenon that is widely reported in most of the nationalism literature (Deutsch, 1953; Hroch, 1985; Hobsbawm, 1990).

[5] Children's names are important cultural markers that disclose parents' racial, ethnic, and social preferences and backgrounds (Fouka, 2019).

during the sample years, other languages showing no significant effects. Further, in the Cape of Good Hope Colony white settler grassroots organizations' meetings that support British institutions increase by 32% while the election of MPs with non-English last names grows by 20%, highlighting how national identity may fuel collective action.

The main identification assumption in this analysis is that in the absence of local increases in market access, national identity choice changes would have followed the same trend everywhere after partialling-out location and time fixed effects. I perform different tests to assess the plausibility of this assumption, addressing different potential threats to identification. First, I examine the possibility infrastructure placement endogeneity by showing that the effects of market access are robust to controlling for local railroad and road placement, a specification that only accounts for variation in distant infrastructure network changes that are not correlated with local identity trends. I control for different distance buffers; coefficients' significance smoothly decreases as distance grows.

Second, I construct a yearly "placebo" infrastructure network using lines that were built *after* the end of the sample years to reach mines that had yet not been discovered. The assumption is that these lines would have been built as early as possible had the mines' existence been known. I then include these "placebos" in the main specifications and show that they do not have a spurious impact on national identity; this shows that line placement decisions are likely not driven by unobservable determinants of identity. Future versions of this project will address the possibility of self-selection into the treatment—or endogenous migration—by restricting the samples to identifiable locals; and examine pre-treatment parallel trends in potential outcomes.

In additional analyses, I then explore the mechanisms behind these results. Following a social identity model, I hypothesize that white settlers' social utility from identifying with a given nation is a function of two components that regulate their acceptance of normative racial integration (Shayo, 2009, 2020). The first is "cognitive distance", which measures an individual's perceived distance to the prototypical cultural attributes of people with whom they share a social category or role: the higher the distance the less utility a person gains from identifying to it. The second is their social category or role's "status" relative to others, which reflects its social value and is positively related to utility. Most white settlers initially identify with the "white farmer" social category.

As increases in market integration may differentially impact local desegregation, changes in white settlers' cognitive distance and relative status may differ across locations—driving divergence in national identity choice. Market integration to non-white populations may expand the local "non-white farmer" pool, increasing white settlers' cognitive costs of identifying as "white farmers"; it may further decrease the racial wage gap as employers replace white- with cheaper non-white workers, decreasing "white farmers" relative status. Similarly, market integration to "non-white farmer" districts may crowd-out "white farmer" agricultural production. Conversely, market integration to white populations may hamper desegregation, leaving white settlers' cognitive distance and status unchanged.

I construct different variables to explore these mechanisms. To capture increases in cognitive distance, I first build a "market access" variable that calculates a locations' costs of trade to white and non-white baseline populations. I then further split baseline white populations into members of the Dutch Reformed Church (DRC)—white settlers' religion

of choice—and other Protestant churches—prototypical incoming British' religions—to test whether belief system, not just race, may also account for the activation of a boundary between British and Afrikaner. Second, to capture variation in white settlers' relative material status with respect to non-white populations, I build variables such as racial wage gaps of different occupations and agricultural production by racial profiles.

Results show that increases in market access to non-white populations induces white settlers to identify more Afrikaner and less British. Market access to white populations has the opposite effect. Belief system salience does not activate a within-white cultural boundary. Further, increases in market access to non-white populations decrease the racial wage gap; increases in market access to white populations increase it. Consistently, increases in market access in non-white agricultural districts crowd-out white production. Desegregation thus increases the marginal value of institutionalized non-white populations' repression, triggering Afrikaner identification. The lack thereof increases the cost of deviating from the British, triggering British identification.

This project contributes to the identity literature in economics, which has highlighted the significant impacts of identity—and culture—on various economic and political outcomes (Akerlof & Kranton, 2000, 2002, 2010; Fernández, 2010; Gorodnichenko & Roland, 2011), as well as its role in shaping institutions (Bisin & Verdier, 2001; Alesina & Giuliano, 2015; Bisin & Verdier, 2017). A growing empirical literature has further studied the determinants of identity, focusing on the impact of states' nation building tools or policies (Dominic Rohner, 2013; Cantoni et al., 2017; Dell & Querubin, 2017; Fouka, 2019; Blouin & Mukand, 2019; Depetris-Chauvin et al., 2020; Fouka, 2020; Assouad, 2021; Atkin et al., 2021; Dehdari &

Gehring, 2022; Algan et al., 2022).[6]

This project adds to this literature in three ways. First, it highlights the role of market integration in shaping people's national identity choice, which provides an approach to analyze the rise of powerful nations with strong normative and institutional preferences when the state itself is non-existent, developing, or weak—scenarios that are prevalent both in history and today (Barth, 1969; Hoetink, 1975; Nagel, 1995). Further, by stressing that nations and state need not necessarily coincide, it underscores how organic or decentralized changes in the economic and social environment may impact social group dynamics in parallel to existing states, leading to the development of nationalisms that can pose threats to even the strongest of sovereign states (Nagel & Olzak, 1982; Munashinghe, 2018).[7]

Second, this project explicitly analyzes the social environment's significance for national identification (Shayo, 2009, 2020). Individuals do not only choose identity within society—as opposed to in isolation—but internalize their social groups' social utilities and cultural boundaries, which allows them to socially interact with millions of people (Barth, 1969; Harari, 2015). In this light, this project changes focus by analyzing national identification as the result of a contest between social groups with different social incentives ("non-white", "(non-)repressive white") for a broader membership base and greater influence in the state (Deutsch, 1953; Weber, 1976; Gellner, 1983; Anderson, 1983; Hroch, 1985; Hobsbawm, 1990), not particular "regimes or constitutions' (Tilly, 1986; Tarrow, 1994).

[6]This literature has analyzed a variety of state interventions such as propaganda, education, military campaigns, charismatic leaders, discriminatory policies, state repression, among various others.

[7]Many examples of such dynamics are currently ongoing all around the world, both in the developed and the developing world, as national, sub-national, or ethnic people's identification empowers social groups outside sovereign states.

Third, this project contributes to the literature by exploring the comprehensive presence of nations in people's private and public lives—from influencing choice of children's first names, the language of print, organization of grassroots meetings, and the election of state representatives—to explicitly highlight the exceptional power of culture. This network of artificial instincts, created and reinforced by powerful myths (Harari, 2015), is capable of aligning heterogeneous individuals into congruent anonymous social networks to effectively regulate social interactions. Successful nations' choice of cultural rules thus permeate all levels of the social space; finding their way into the home, the world by means of collective action and, finally, the official institutions of the state (Hroch, 1985).

I borrow the "market access" measure from a long-standing literature that analyzes the impacts of infrastructure expansion. The economic effects of railways was first studied by Fogel (1964) and Fishlow (1965) and has been extensively researched ever since (Donaldson & Hornbeck, 2016; Perlman, 2016; Katz, 2018; Hornbeck & Rotemberg, 2019). Other studies have followed suit by analyzing the impact of railways on urban population growth, spatial population structure, and structural change around the world (Alvarez et al., 2013; Berger & Enflo, 2014; Hornung, 2015; Jedwab et al., 2015; Bogart et al., 2018; Donaldson, 2018; Banerjee et al., 2020). This project builds on these geographical tools to analyze its impacts on the social space: identity, culture, and normative and institutional preferences.

This project also resonates with the literature that analyzes the diffusion of behaviors, much of which focuses on online social networks (Bakshy et al., 2012; Aral & Walker, 2012). Studying networks in a historical setting provides some advantages. While online networks face an empirical limitation due to links being defined by people's selection into friendships,

historical cases allow studying networks that were costly to build and had clear economic objectives, which allows designing strategies to address identification concerns such as placebo networks. Studying historical networks also contributes to the literature that studies the impacts of face-to-face interactions on innovation, markets for ideas (Andersson et al., 2020), and search and contracting costs (Startz, 2017).

Finally, this project relates to the literature that explores the effect of technological shocks on political mobilization, the findings of which have been mixed. One set of papers finds that technological shocks facilitate grassroots political coordination and participation by increasing social interactions (García-Jimeno et al., 2018; Melander, 2020). A second group shows that technology decreases protest activity when it leads to improved employment opportunities and bargaining positions (Falck et al., 2014; Molinder et al., 2019; Caprettini & Voth, 2020). This project contributes to both sides of this literature by analyzing one technological shock—infrastructure expansion—with diverging effects on national identity choice, one nation more grassroots the other more top-down.

This document is organized as follows. Section 2 provides a historical overview of the spread of British and Afrikaner identity in colonial South Africa. Section 3 presents a conceptual framework. Section 4 describes the data collection and variable construction. Section 5 discusses the empirical strategy. Section 6 presents results. Section 7 analyzes the mechanisms predicted by the social identity model. Section 8 concludes.

2 Historical Overview

White populations are dominant in 19th century colonial South Africa and racial divisions are largely impermeable;[1] they have developed different norms and institutions due to two separate waves of colonization and settlement. The first wave are descendants of Dutch East India Company (VOC) workers that first arrived in the 17th century. Most are farmers that live in near-autarky; their immediate surroundings define their social identity (Davenport, 1960). They rely on coercive labor and the enslavement of non-white populations for domestic and farm work, however outlawed (Morton, 1992).[2] These settlers also control local (and underdeveloped) labor and capital markets through particularistic and personal networks (Nagel & Olzak, 1982; Olzak, 1992).

The second wave settled in ports and towns in the early 19th century to pursue commercial activities: the British. They have a strong social identity backed by a global Empire centered in London, a standardized language, and unifying foundational myths.[3] The British central state and private companies run and control the widest trading markets in the world (Beckert, 2015). While the ruling class is white, by the 1870s slavery has been replaced with free wage labor and non-white males are enfranchised in one South African colony. British and rural white settlers have limited contact and overall communication is fractured; not only are they physically separated, but farmers speak mutually unintelligible local dialects, are typically illiterate, and rarely understand English (Davenport, 1960).

[1] As described in Toynbee's (1959) characterization of British and Dutch colonial models.

[2] With institutions such as *Inboekstelsel*, or bonded labor from stolen or purchased children. Boers also organized raids to capture African slaves for years after the different states had prohibited such practices.

[3] For a brilliant (and fun) exposition of British foundational myths see Chesterton (1917).

The status quo is disrupted when the world's largest diamond (1871) and gold (1876) fields are discovered in the rural interior. The mineral revolution galvanizes the expansion of a massive transportation network that decreases transportation costs and deepens market integration, ending white settlers' autarky by pressuring norms and institutions that govern the isolated interior to adapt to the global economy (Henrich, 2020). As new economic opportunities arise for all racial groups, the long-established racial hierarchies of the rural interior erode (Nagel & Olzak, 1982; Olzak, 1992). The tension between the economic benefits of integration and the dissolution of traditional social structures is at the center of white settlers' identity choice under the forming nation-state (Deutsch, 1953).

White settlers can either identify with the well-established British or can adhere to the nascent Afrikaner. The choice implies adopting each social groups' norms and institutions. Joining the British social group involves sharing the social, economic, and political space with both white and non-white populations. The Afrikaner nation promises to narrow such interactions by activating an ascriptive boundary (by race, language, religion, and shared narratives) that ties all members under a sole social group ensuring separateness or *apartheid* (Barth, 1969). Noticeably, identifying with the Afrikaner implies actively rejecting all things British: this includes the English language, religions, and liberal market-oriented economic paradigms.

This project analyzes white settlers' identity choice during the formative years of the South African nation-state: from the start of the increase of market integration in 1869 to the official creation of the Union of South Africa in 1910.

3 Conceptual Framework

I conceptually discuss how increases market integration resulting from decreases in transportation costs may affect white settlers' identity choice. The argument is as follows. First, I hypothesize that as traditional social structures erode, large-scale social groups that coordinate people's access to social categories and roles using cultural criteria are most fruitful at inducing mass identification. Nations are one such group. Transversal disagreements regarding the usage of such criteria in large-scale society—such restricting access based on somatic, religious, or kin requirements—may however activate a within-white cultural boundary that reflects this dissent. White settlers' identification divergence across Afrikaner and British may indicate that such a disagreement exists.

Second, I hypothesize that race is at the core of this transversal disagreement. In particular, market integration's differential impact on local desegregation may differently affect white settlers' cognitive cost of identifying as "white farmers" as well as the relative status of the "white farmer" social category relative to other "non-white" ones. If desegregation—or racial mobility—increases, white settlers' marginal value of institutionalizing non-white populations' repression may increase, inducing them to identify with the Afrikaner since British institutions disallow repression at a high cost. If desegregation is not a threat, white settlers may identify as British since the marginal cost of deviating from British norms and institutions that dominate world markets increases.

3.a Market Integration, Social Identity Scale, and the Role of Culture

In an initial period, locations (districts or subdistricts) are weakly connected to one another due to high transportation costs. Each location has its own set of local norms and institutions that specify—either formally or informally—the social categories to which local inhabitants belong to and define their social roles (Akerlof & Kranton, 2000, 2002, 2010). In the colonial context, social categories are typically defined by racial criteria: white populations control local politics and labor and capital markets through particularistic and personal networks (Nagel & Olzak, 1982; Olzak, 1992) and rely on non-white populations' coerced labor or enslavement (Morton, 1992). White populations' identify as "white farmers", a social category that reflects their privilege in the local racial hierarchy.

In subsequent periods, the development of infrastructure networks decreases transportation costs and increases local market integration. Locations with higher increases in market integration are differentially penetrated by external large-scale social anonymous networks with different norms and institutions that define access to social categories and roles (Olzak, 1992). White traditional control over local politics and economics erodes giving way to externally determined large-scale organizational structures, such as a state bureaucracy or homogeneous markets that rationalize labor and capital allocation based on productivity instead of tradition. Local power structure shifts change available social categories and, consequentially, white settlers' identities (Hannan, 1979; Nielsen, 1985).

While identities in large-scale societies may take many forms,[1] social groups that align

[1]They can be based on occupation, family or kingship, territorial units, or demographic, biological, and physical characteristics, among infinitely many others (Nielsen, 1985).

interests of large transversal shares of the social anonymous network's population have higher success rates (Akerlof & Kranton, 2010; Henrich, 2020). Structural models of social identity choice suggest that groups that restrict or expand access to social categories and roles using cultural requirements are most fruitful at inducing mass identification due to culture's malleability and overarching applicability to individuals of different occupations, classes, kins, and traits such as gender or age (Hannan, 1979; Nielsen, 1985; Olzak, 1992). Nations are one such *imagined community* (Deutsch, 1953; Weber, 1976; Gellner, 1983; Anderson, 1983; Hroch, 1985; Hobsbawm, 1990; Tilly, 1986; Laitin, 2007).

In the colonial context, white populations may in principle replace local traditional identities with any available large-scale identity (Hannan, 1979). Given time and energy constraints however, identification decisions may be leveraged by how social groups' proposed selection mechanisms that regulate access to economic and political resources map from small- to large-scale societies. The nation that proposes the cultural criteria to coordinate people's access to social categories and roles—and thus resources—that least disrupt such channels may most induce white settlers' mass identification,[2]. More than one nation may however arise if transversal normative disalignments exist regarding such cultural criteria, leading to the activation of a within-white cultural boundary (Barth, 1969).[3]

Evidence of white settlers dividing their identification between Afrikaner and British nations suggests that one such significant transversal disagreement does indeed exist.

[2]Identification does not necessarily imply collective action or mobilization. It is a more passive form of group membership. Identity might be *expressed* through collective action, but it need not necessarily be the case.

[3]Note that individuals may change identification across boundaries as their circumstances change; strong successful social groups (i.e. nations) have boundaries that exist independently of the membership of any one particular individual or small sub-groups.

3.b British or Afrikaner? Choosing a National Identity amid Racial Mobility

One such transversal disagreement may be whether to allow different racial groups to share social categories and roles, as non-white populations economic and political mobility increases—which are tightly intertwined.[4] Market integration might provide non-white populations with tools to break out of traditional coercive labor or slavery by expanding opportunities in the labor market and agricultural production (Bundy, 1979). Further, economic mobility might offer disenfranchised populations a path towards broader political rights, such as access to voting based on income qualifications (Markoff, 1996). While the British normatively accept non-white populations' gradual mobility for the sake of preserving dynamic labor and capital markets, the Afrikaner do not.

Differential increases in market integration may unevenly impact non-white populations' social mobility and, consequentially, white settlers' national identity response. Following a social identity model, I hypothesize that white settlers' utility from identifying with a given nation is a function of two components that regulate their acceptance of normative racial integration (Shayo, 2009, 2020). The first is "cognitive distance", which measures an individual's perceived distance to the prototypical cultural attributes of people with whom they share a social category or role: the higher the distance the less utility a person gains from identifying to it. The second is their social category or role's "status" relative to others, which reflects its social value and is positively related to utility.

Increases in market integration may differentially impact desegregation, leading white

[4]I use "non-white" to refer to African, Colored, and Indian populations in South Africa. This category was broadly used at the start of the sample (1870); only during the turn of the century were more racial categories introduced.

settlers' cognitive distance and relative status to differ across locations. This drives divergence in national identity choice. If the cognitive cost of identifying with traditional "white-farmer" social category increases or "white-farmers" lose status relative to other "non-white" social categories, white settlers' marginal value of institutionalizing non-white populations' repression increases, inducing them to identify with the Afrikaner since the British disallow repression. If desegregation along these dimensions does not occur, white settlers identify as British since the marginal cost of deviating from British norms and institutions that dominate world markets increases. Section A depicts a model.

Changes in cognitive costs and relative status may coincide. Increases in market integration to non-white populations may expand the local "non-white farm" worker pool, increasing white settlers' cognitive costs of identifying as "white farmers"; it may further decrease the racial wage gap as employers may replace high wage white workers with cheaper non-white workers, decreasing the status of the "white farmer" relative to the "non-white farmer". Similarly, market integration increases in non-white agricultural districts may lead productive "non-white farmers" to crowd-out less productive "white farmers" in traditional staples. Conversely, integration to white populations may hamper desegregation, leaving both white settlers' cognitive costs and relative status unchanged or higher.

I finally hypothesize that potential changes in cognitive costs brought about by variation in cultural attributes such as belief systems—namely religions—may not be significant drivers of the divergence in white settlers' national identity choice. Traditional social structures' restriction of people's access to social categories and roles based exclusively on racial criteria may explain this; white populations possess equal rights whatever their beliefs. Variation in

market integration to white populations with different or similar belief systems—however unpleasant or pleasant—may therefore not translate into differential access to social categories and roles as potentially disadvantaged white populations may use their baseline legal rights to counteract such attempts.[5]

Baseline small-scale traditional social structures' norms and institutions may therefore influence successful large-scale nations' preferred norms and institutions. While belief system criteria may be used in other contexts by other nations to restrict or expand access to social categories and roles, in colonial South Africa large-scale nations' norms and institutions inherit small-scale local traditional social structures' concern with race. White British and Afrikaner populations remain equal under either nation's norms and institutions; whichever white group controls the state, however, determines whether or not non-white populations are equal as well.

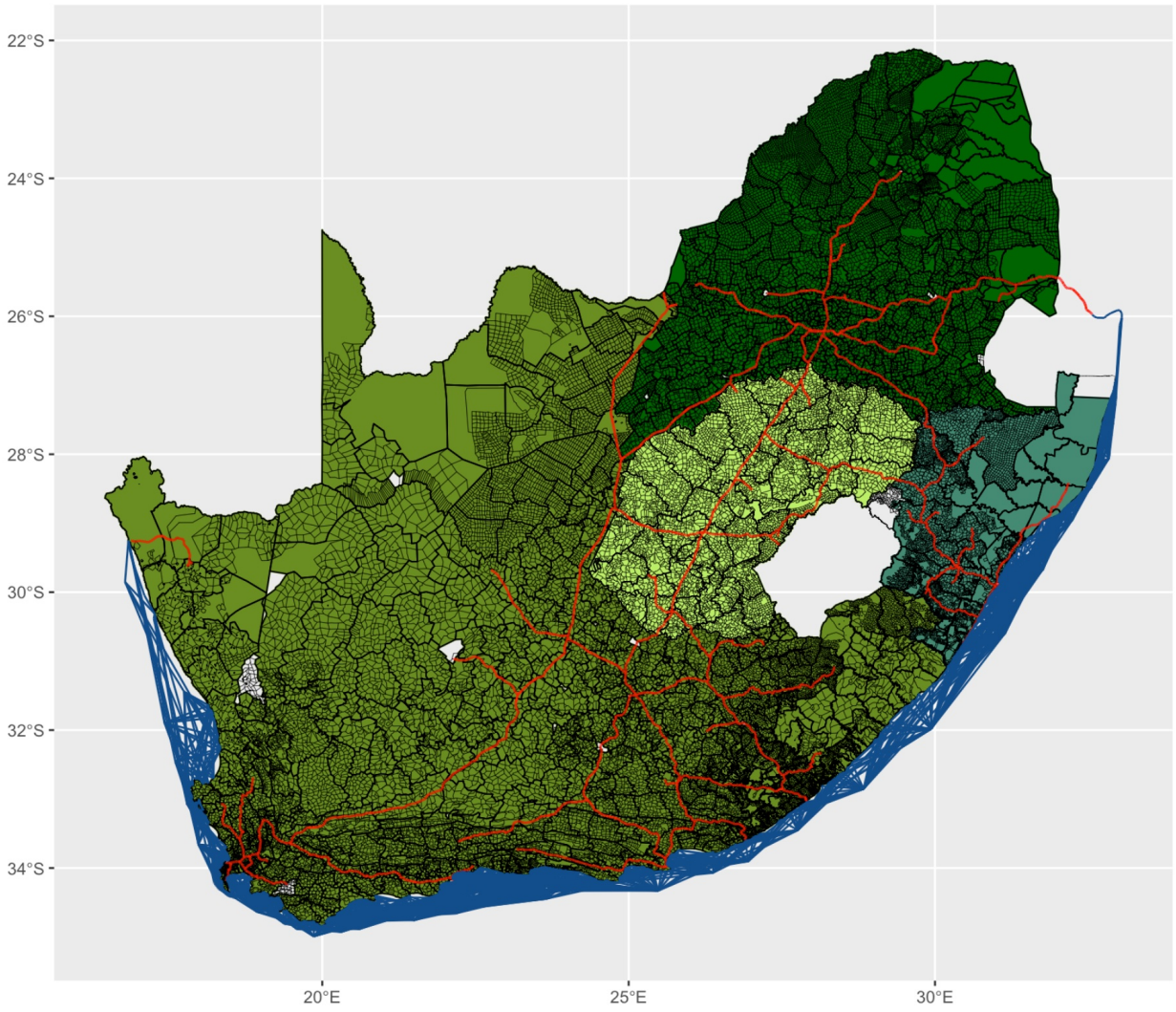
[5]The Transvaal (ZAR) attempted to restrict white *uitlanders* or foreigners' access to voting. This restriction did not last long.

4 Data

I build a novel dataset using a variety of primary and secondary sources and integrate it into one congruent geographical spatial object. I created the baseline spatial object by unearthing hundreds of local land surveys and developing a technology to georeference 19th century triangulation projections into current digital coordinate systems. The result is depicted in Figure 1.[1] I then merge three main sets of variables onto the spatial object. First, a set of aggregate market access variables to measure market integration (Donaldson & Hornbeck, 2016). Second, variables that capture private and public expressions of national identity. Third, economic variables collected from censuses and statistical books to explore mechanisms. I describe the construction of these variables in detail below.

[1]These surveys were collected in the Cape, Bloemfontein, Pretoria, and Pietermaritzburg archival repositories in South Africa, Kew Gardens in Britain, and the Berlin Lichterfelde Bundesarchiv in Germany. I also consulted libraries in South Africa, Germany, Britain, Australia, and the United States to access some of them. See the references at the end for details.

Figure 1: Georeferenced Spatial Object of South Africa



This figure depicts the baseline spatial object that ties this project's data together. It gathers information from the four different polities that constituted what is today South Africa between 1870 and 1910: the Cape of Good Hope Colony, the Colony of Natal, the Orange Free State (OVS), and the Transvaal (ZAR). Smaller polygons are farms, intermediate polygons are field-cornetcies or sub-districts, and districts are made of several sub-districts (not explicitly colored for clarity). The red lines are railways in 1890. The blue lines are sea routes. To build this object, I developed an original technology that allowed me to georeference hundreds of historical land surveys at a very detailed level (see Figure **??** for an example of an unprocessed survey map). These surveys contain information on geo-political boundaries, transportation infrastructure networks, and historical locations that do not exist today, providing the cornerstone that allows to tie together church records, newspaper publications, grassroots meetings advertisements, elected Members of Parliament, censuses, statistical reports together into a congruent corpus of data. Other data such as voter rolls and land tenancy—which I will analyze in subsequent projects—may also be merged into this spatial object.

4.a Market Access

I built a set of aggregate market access variables closely following the methodology developed by Donaldson and Hornbeck (2016). I describe three main steps below.

4.a.1 Defining Transportation Cost Parameters

I first define transportation cost parameters using calculations by Fourie and Herranz-Loncan (2015). They estimate an average rail freight cost of £0.0065 per ton-mile while the ox-wagon transport rate is £0.021 per ton-mile (calculated using the prices of barley, maize, oat hay, and wheat) and sea routes rates are £0.014 per ton-mile. Given that the survey maps differentiate by road type I define the road freight cost price as infinitesimally higher or lower depending on the quality of the road. Table 1 shows the different types of roads included in this project's computations (grand trunk, trunk, main, etc.). All calculations therefore take into account the cheapest routes from location i to location j conditional on the existing rail and road network.

Furthermore, as will be described in more detail below, roads differ from railroads as they are constantly rebuilt with the passing years. Primary sources reveal that roads have a high depreciation rate, making them intransitable relatively fast: even the highest quality "Grand Trunk" roads had to be repaired several times during the sample years. I therefore include in the parameters a depreciation rate δ that decreases the quality of the road on a yearly basis if it is not repaired in the given year. From contemporaneous engineering reports I set $(1-\delta) = 0.875$ depreciation per year.[2]

[2]This number was taken from statistical books. See the primary sources list below.

Table 1: GIS Network (1869-1909)

Type	Cost freight (ton-mile)	Data Source	Description
		Railroads	
All lines	£0.0065	Statistical registers, land surveys, and secondary literature	Hand-traced georeferenciation and shapefiles
		Roads	
Grand Trunk	£0.021	Statistical registers, and land surveys	Hand-traced georeferenciation
Trunk	$£0.021 + \varepsilon^{+}$	"	"
Main	$£0.021 + \varepsilon^{++}$	"	"
Branch	$£0.021 + \varepsilon^{+++}$	"	"
Farm	$£0.021 + \varepsilon^{++++}$	Linear projection	Direct route to nearest railroad or road
Bridges	£0.021	Statistical registers, and land surveys	Hand-traced georeferenciation
Passes	£0.021	"	"
		Sea Routes	
All lines	£0.014	Ports in 1870-1910	Hand-traced georeferenciation

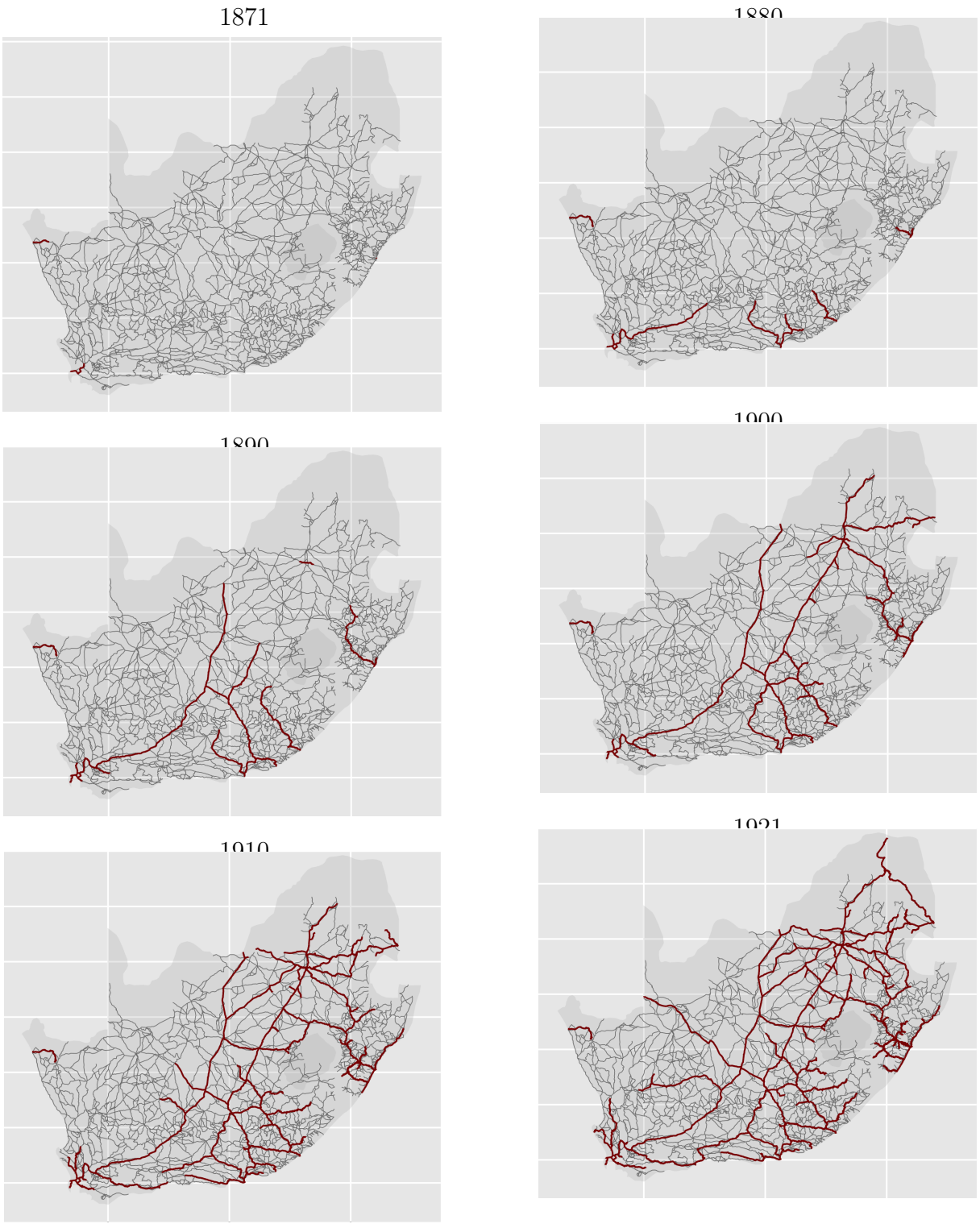
This table details for each of the different types of transportation infrastructure built into the spatial object's costs of use, its datasource, and the description of how it was georeferenced. Railroads and sea routes all have the same transportation costs. Roads however vary in quality. Highest quality roads are Grand Trunk, followed by Trunk, Main, Branch, and Farm. I assign the highest quality to Bridges and Passes since they are constantly being renovated as they are essential for transportation. I built all transportation infrastructure information using a mix of statistical registers that detail the years in which construction takes place and the georeferenced spatial object that locates this construction in space. All network lines were built by hand (for railroads I used current PlanetGIS networks as a guide as well).

4.a.2 Transportation GIS Network

To build the network of railroads and roads I first used systematic government statistical reports to add onto the georeferenced land surveys described above and depicted in Figure 1. The existing documentation allowed me to rebuild at the very detailed geographical level the yearly construction of railroad lines between 1869 and 1909. Using PlanetGIS contemporary shapefiles that depict current open and old closed railroad lines as a guide to locate original railroad routes, I was able to reconstruct the year-by-year historical network using historical reports that describe these railroad openings by year. The output of this work is depicted in Figure 2. I build the sea routes by connecting all the existing ports to each other with a saturation of straight lines.

Finally, I construct the road network in a slightly different way. Detailed road building records were published yearly in Blue Books and Departmental Reports, which precisely describe the locations of roads, bridges, and mountain passes and the type, duration, and even the total cost of the work being performed. To build the panel I first drew a road network using land surveys around 1901 for the four territories linked to the spatial object, as guide for road placement. I then transcribed the statistical registers, matched the historical locations to the georeferenced spatial object, and drew the yearly construction of roads using the guide grid. Figure B1 shows a heat map that aggregates all the roads built between 1869 and 1909 for the whole territory.

Figure 2: Railroad Network Expansion (1869-1909)



This figure shows in red the expansion of the railroad infrastructure for different years between 1869 and 1921. Railroads are virtually non-existing in 1870; by 1910 they have expanded considerably.

This heat map reveals strong regional differences in infrastructure development and state-capacity. First, clearly shows that some specific routes and regions are repaired many more times than others, where those connecting ports and mines seem to get predilection. Second, the British Cape of Good Hope Colony overwhelmingly leads in road building compared to the other British Colony of Natal, and the Orange Free State (OVS) and Transvaal (ZAR) territories. This hints to stronger state-capacity. To highlight the topologies' precision level, the left panel of Figure B2 shows an example of an intersection between roads and rails. Further, the right panel shows the expansion of telegraphic lines highlighting some years, which I analyze in subsequent versions of this project.

The next step was to process and merge the "hand-traced" topologies (see Table 1). I used R-Grass to break, snap, and clean the railway and road raw edge geometries. The result is a smooth coordinate vector space that allows any agent or good to travel through the network and change direction at each intersection (see Figure B2). A crucial advantage of using such CLI geographical software to analyze network structures is that it is extremely flexible. Not only is it easy to assign various transportation cost parameters (τ) to each technology type, but it is also straightforward to set additional costs to hypothetical agents that switch transportation technologies as they move through the network—such as those incurred in when passing through different quality roads.

I then import this vector space into R-iGraph to analyze the data. I connect the 1,282 locations to the railway and road network in each of the sample years. I do this by creating pseudo-edges that link each location to the closest part of the network in a given year by creating a projection; I assign to these pseudo-edges the transportation cost parameters

of the lowest quality roads (see Table 1). The left panel of Figure 3 depicts the abstract representation of such a graph for one particular year. The graph is detailed: it is composed of hundreds of thousands of edges and nodes all of which are connected to one another. These edges hold all transportation cost information needed to calculate the market access variables.

The resulting network therefore connects a set of locations $i = 1, 2, ..., n$ by railway, road, and sea route. Each year location i's connectivity to the network varies with changes in the construction of the full network of infrastructure. With this setup, I create a $n \times n$ ($n = 1,282$) matrix in which each element is the transportation cost τ_{ij} between location i and location j calculated using the shortest path Dijkstra algorithm.[3] I then repeat this calculation for each year in the sample; all locations are included in every year's calculations (Donaldson & Hornbeck, 2016). The right panel of figure 3 shows *one* shortest path between two random points in 1890 for illustration (this route starts by road, then railway, and ends by road again).

4.a.3 Computing Market Access

The final step is to calculate the market access variable for each location i. This variable may be represented as follows (Donaldson & Hornbeck, 2016):

$$MA_i \approx \sum_{j \neq i} \tau_{ij}^{-\theta} N_j$$

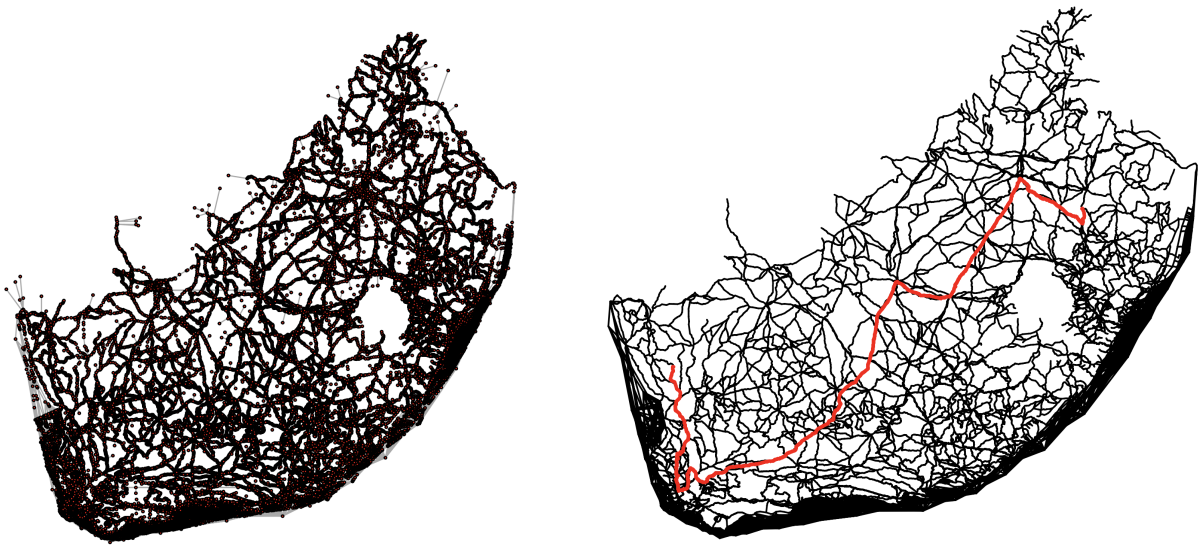
Where i is a location, $j \neq i$ are all other location in the sample, τ_{ij} are the transportation

[3]This is the most standardized algorithm to do this calculation. It is also the most common one used in the literature (Donaldson & Hornbeck, 2016; Donaldson, 2018).

Figure 3: iGraph and Dijkstra Path Calculations

(a) iGraph Abstract Representation

(b) Dijkstra Calculation Example



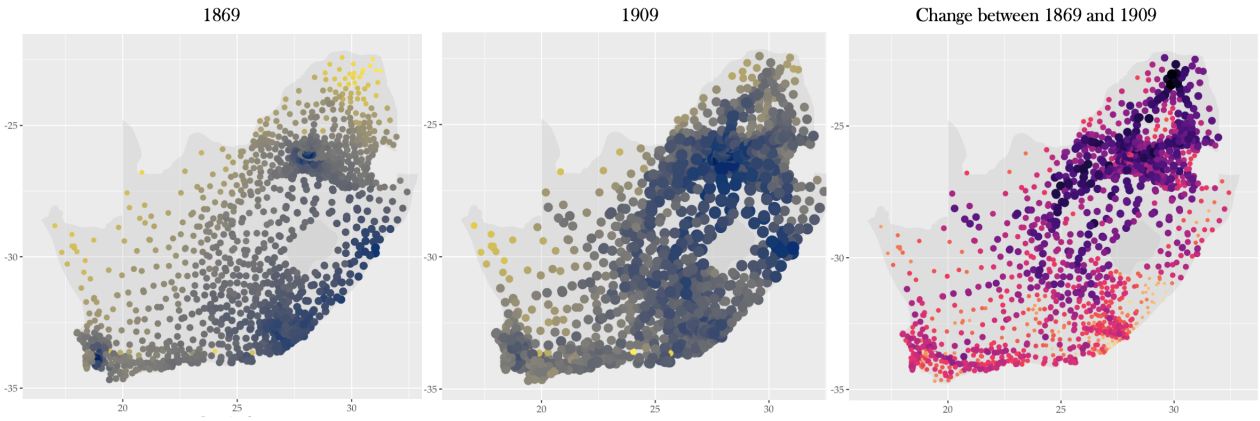
The left panel shows the abstract iGraph with points of interest (POIS) connected to the network. The graph holds all the geographical and cost parameters information needed to calculate the Dijkstra Shortest Path between any two points in the geographical space. The right panel shows one example of a georeferenced shortest path calculation between two random points using the Dikjstra algorithm.

costs between locations i and j, θ is "trade elasticity" parameter, and N_j is number of people in location j. N_j is the population of location j in the baseline year, 1870. Crucially, for the purposes of studying mechanisms I calculate three different market access variables: where N_j can be either the total, the white, or the non-white population in location j. As I describe in more detail below, the population information was collected from censuses of the four historical polities that constituted today South Africa (Cape, Natal, OVS, and ZAR). Note that location i's own N_i is always excluded from the term MA_i in order to reduce endogeneity concerns.

The θ parameter has a direct computational impact on the MA_i variable. I construct

MA_i choosing several θ; this document presents results with $\theta = 1$. Yearly MA increases monotonically as expected: Figure 4 depicts geographically these changes in MA_i. The 1,282 locations are included in all three panels. The first panel shows MA_i in 1869; the second panel shows MA_i in 1909, and the third panel shows the change between both years. Darker colors mean higher MA_i. The visuals are straightforward: market access in i increases the more railroad and road networks expand. Noticeably, the regions of the interior—away from the coasts and closer to the mines—experience a higher increase in market access during the sample years.

Figure 4: Market Access Variable in 1869, 1909, and Change Between Years



The first panel shows calculated market access in 1869; the second panel shows market access in 1909, and the third panel shows the change between both years. Darker colors mean higher market access. The visuals are straightforward: market access in any given location increases the more railroad and road networks expand throughout the territory. Noticeably, the regions of the interior—away from the coasts and closer to the mines—experience a higher increase in market access during the sample years.

4.b National Identity Variables

This section describes the construction of the various national identity variables. The measures reflect the grassroots origins of Afrikaner identity; they highlight the standardization of a

novel language from oral common-usage vernaculars and pidgins, the surge of bottom-up political and economic civil organizations, and the institutionalization of a new political elite to legitimize the Afrikaner national agenda. Conversely, the variables also capture the picture of a growing British national identity that is being introduced in a top-down fashion by an organized, strong, and fully formed expanding state; they evidence the spread of the English language, the enhancement of British-oriented grassroots organizations, and the continued grasp of English politicians in the state.

Moreover, these variables reflect the wide scope that is typical of national identities: they not only permeate people's private lives in the form of parents' choice of children's first names, but they also impact the public arena through public discussion that can only become widespread with the development of a standardized language, shape the nature of civil organizations, and—when successful—its ideals, narratives, and myths can find their way into the state itself (Deutsch, 1953; Anderson, 1983; Hobsbawm, 1990). The dynamic interplay between popular mobilization and elite leadership is central in shaping successful nations (Hroch, 1985). Nations' overarching reach accounts for what is perhaps one of the most successful social identities in recent history.

4.b.1 Children's First Name Etymologies

Childrens' names are markers for parent's identity affiliation (Fouka, 2019, 2020). In order to capture the first name's etymologies I collect birth and baptism records from the Dutch Reformed Church (DRC) available in genealogical websites.[4] This datasource is ideal to

[4]These sources were collected from *Family Search*. There are four main archives, all included in this project: the Free State, Pretoria, Cape, and Natal.

capture individuals' adoption of British or Afrikaner national identity for two reasons. First, churches were the institution responsible for the registration of births, baptisms, and marriages up until the first quarter of the 20th century in South Africa, making them a key source of information. Second, the DRC was the church of choice of the white settler base that is deciding whether to identify with the incoming and dominant British or the nascent Afrikaner national groups (Davenport, 1960; Gilliomee, 2009).[5]

Table 2: Afrikaans Naming Rules (Examples)

Non-Afrikaans	**Afrikaans**	**Change**	**Linguistic Rule**
Carl	Karel	c → k	Assimilation
Rensche	Rentia	sche → tia	Dissimilation
Diederick	Diederik	ick → ik	Haplology
Machteld	Magtel	ch → g	Simplification
Hermina	Hermiena	mina → miena	Metathesis
Henriëtta	Henrietta	ë → e	*umlaut loss*

This table shows some examples of changes in non-Afrikaans names—such as Dutch, French, German, Khoisan, and Malay—after applying the linguistic function that captures the consolidation of Afrikaans naming practices. Afrikaans is a créole language that organically evolved from a variety of pidgins and vernaculars, or *Kitchen Dutches*, eventually becoming a new standardized language. The linguistic function therefore marks the evolution of the language from common day-to-day oral usage into a standardized orthography.

The yearly DRC records comprises of 374,330 births and baptisms between 1870 and 1910.[6] This is an arguably representative sample. The 1911 Union of South Africa census counted 1,276,242 Europeans (or whites, from a total of 5,973,394 persons) of which 693,898 were members of the DRC: the records therefore account for births of 54% of all people registered as DRC members (alive members absent in the records include those born before

[5]There are several accounts of the DRC not allowing non-white members. African, Colored, and English populations were typically affiliated with Wesleyan Methodists, Anglican Communion, among others (Davenport, 1960).

[6]When birth information is unavailable I choose the baptism date instead; children would be baptized not long after their birth.

1870, or older than 41 years in 1911). I first merge the birth locations reported in the DRC records—several of which do not exist today—to the spatial object described above (see Figure 1). I was able to match this data to 687 urban and rural locations accounting for 96% of the full church records sample.

I then build a novel database to measure the consolidation of Afrikaans naming practices. Afrikaans is a créole language that organically evolved from a variety of pidgins and vernaculars of Dutch, French, German, Khoisan, and Malay (a.k.a. "influence languages"), or *Kitchen Dutches* (Deumert, 2004). This project's linguist created a function that marks the evolution of the language from common day-to-day oral usage into a standardized orthography.[7] These changes highlight the bottom-up origins of the new national identity's language: phonological simplifications, transpositions, and elisions become written rule (Krige, 1936; Heese et al., 1975; Raidt, 1985; Jansen et al., 2007; Heeringa & DeWet, 2015). Table 2 shows examples of the formal linguistic function.

Third, I identify children first names' etymology. In the context of the sample years (1870-1910), white settlers have a *menu* of identities to choose from: Afrikaans, English, Dutch, or other European. To identify the "influence language" first names I use online etymological databases.[8] To identify the Afrikaans names I apply the linguistic function to "influence language" names, coding them as Afrikaans if they comply with the linguistic rule and are different to any "influence name". Since one child can have up to five first names and one name can share several etymologies, I code each child's name etymology *without replacement*

[7]Chané Venter is this project's main linguist. She created the linguistic rules function. The guidance from professors Wannie Carstens and Rajend Mesthrie was instrumental to the development of this function.

[8]I use *Behind the Name*, *Names.org*, and *House of Names* online name repositories.

using the following rule: Dutch if they have at least one name that is etymologically Dutch, English if at least one is English, and finally Afrikaans.

This definition errs on the side of caution. Since Dutch is the most common "influence language" among Dutch Reformed Church members, I code the child as Dutch if at least one name is etymologically Dutch, even if the Dutch name might also share an English etymology or if the child has any other English or Afrikaans names. This means that the children that are coded as English do not have any Dutch first names at all, and those that are coded as Afrikaans do not have any Dutch or English first names, making these children *extremely* Afrikaans. Table 3 shows the ten most common Afrikaans, English, and Dutch names using this rule.

This table confirms that Dutch and English are more standardized languages than Afrikaans, as the first ten names account for 42% and 35%, respectively, of all the children's names of each etymology, compared to 20% of Afrikaans—meaning that most names are still uncommon. Table 4 summarizes the means and standard deviations for each variable when aggregated by sub-district, the level of analysis. The yearly average sub-district share of children with Afrikaans last names is lowest (1%), followed by English (7%) and finally Dutch (91%). This is consistent with the historical context, where the nascent Afrikaans is slowly gaining momentum in white settler society amid standardized languages such as the incoming English and the baseline traditional Dutch.

Table 3: Ten Common Children's Names (1870-1910)

Name	Absolute Frequency	Relative Frequency	Cumulative Relative Frequency	Ranking
		Dutch		
Johannes	50766	7.85	7.85	1
Maria	38035	5.88	13.73	2
Johanna	36526	5.65	19.38	3
Jacobus	32130	4.97	24.35	4
Catharina	22192	3.43	27.78	5
Petrus	19425	3.00	30.78	6
Anna	19114	2.95	33.73	7
Susanna	18795	2.90	36.63	8
Magdalena	18660	2.88	39.51	9
Hendrik	18095	2.80	42.31	10
		English		
Elizabeth	2663	8.25	8.25	1
Hester	1413	4.38	12.63	2
Daniel	1325	4.10	16.73	3
William	1084	3.36	20.09	4
Sophia	949	2.94	23.03	5
George	889	2.75	25.78	6
Francois	799	2.47	28.25	7
James	747	2.31	30.56	8
Henry	685	2.12	32.68	9
Charles	619	1.92	34.60	10
		Afrikaans		
Margrietha	101	3.14	3.14	1
Johannis	80	2.48	5.62	2
Pieternella	77	2.39	8.01	3
Elizabetha	64	1.99	10.00	4
Lodewikus	63	1.96	11.96	5
Janse	61	1.89	13.85	6
Michal	55	1.71	15.56	7
Izabella	47	1.46	17.02	8
Jansen	44	1.37	18.39	9
Lodewyk	44	1.37	19.76	10

This table shows the ten most frequent names of each etymology. Dutch and English are more standardized languages than Afrikaans, as the first ten names account for 42% and 35%, respectively, of all the children's names of each etymology, compared to 20% of Afrikaans—meaning that most names are still uncommon.

4.b.2 Newspaper Language

Print languages are an expression of people's national identities (Anderson, 1983). Common language aligns individuals from different different backgrounds into a shared communication space. I use a State Library compilation that indexes every known newspaper published in South Africa to access this information (Aschenborn, 1983). This source highlights the richness of publications. Print comes in a plethora of languages, focuses on topics from the regional to the international, and editors, reporters, and contributors reflect the diversity of the South African colonial society itself. I extracted details on the locations, years, and frequencies of these published periodicals. I track hundreds publications in 131 locations between 1870 to 1910.

In order to measure the language of newspaper publications I again use language markers. Accounting for English newspapers is straightforward since the language is already standardized by 1870. On the other hand, Afrikaans starts to noticeably deviate in print from Dutch and the other "influence languages" as early as the 1860s.[9] Given the language's grassroots evolution, the change into its current form is not immediate and goes through different phases which are accounted for in the creation of this variable—since its variations are noticeable to linguists' eyes. The organic evolution of Afrikaans is the more apparent when considering it is only granted formal language status in 1925, many years after this project's last sample year (Gilliomee, 2009).

[9]The first Afrikaans book is considered to be *Zamenspraak tusschen Klaas Waarzegger en Jan Twyfelaar* (*Conversation between Claus Truthsayer and John Doubter*) by L.H. Meurant published in 1861. It is however noteworthy to mention that the first known written form of Afrikaans appeared in the early 19th century in Cape Muslim schools in Arabic script (Gilliomee, 2009).

Table 4 summarizes the means and standard deviations for each newspaper language when aggregated by district, the level of analysis. The yearly average district number of newspapers between 1870 and 1910 in Afrikaans is 0.2, English is 0.37, African and Indian languages is 0.02. This is consistent with the historical context where, for example, the maximum number of newspaper publications per year in any given city is 19 (Johannesburg), with a total 2,340 total issues printed that year. Most newspapers are ephemeral, since their main income are subscribers and—as contemporary reports state—most fail to keep a constant reader base. Some newspapers however exist for several decades, such as *De Zuid-Afrikaan* which is published from 1830 to 1930.

4.b.3 Grassroots Organization Meetings (Cape of Good Hope Colony)

Grassroots organizations' agendas reflect members' preferences and—importantly—the social categories they identify with. The last quarter of the nineteenth century experienced a massive increase of grassroots organizations in the Cape of Good Hope Colony that addressed a variety of different topics, evidence of a strong developing civil society. White settlers created organizations to support different common interests and causes that ranged from the regional to the national and from the economic to the political. This organic rise of grassroots organizations reflects both the development of the Afrikaner nation, on one hand, and white settlers' alignment to British norms and institutions, on the other. Two organizations are of particular interest.

The first one is the *Boeren Beschermings Vereeniging*'s (Farmers Protection Association) whose origins can be traced to local associations that independently organized in the East,

Midlands, and West of the Colony during the 1870s at the hand of non-English speaking farmers that sought to further their economic interests as markets permeated the interior (Davenport, 1960, p. 37). Second, a more extreme religious, linguistic, and racially oriented fringe organization that spread from the Western Cape to the rest of the Cape Colony and beyond: the *Afrikaner Bond*. The Bond became the first official political party of South Africa, being particularly successful at channeling the social sentiments of white settlers in marginalized rural areas (Davenport, 1960).

The relationship between these two organizations was tense from the start. The traditional segregationist values of Bond members were in direct contradiction with the Cape Colony's "color blind" institutions the Farmers Association originally affiliated with; the strict preservation of racial hierarchies across social, political, and economic spheres clashed with economic concerns (Davenport, 1960; Grundlingh, 1945, p. 230, p. 232). These opposing positions highlight again the fundamental tensions between British and the nascent Afrikaner nation regarding the topic of race in South Africa. As MP Florence Fuller bemoaned: "God help the country [...] if we are about to enter on a crusade against the rights and privileges enjoyed by the Natives [...]" (McCracken, 1967, p. 51).

In order to trace the gradual formation of these grassroots organizations I collected forty years (1870-1910) of the *De Zuid-Afrikaan* newspaper issues that advertised every meeting held by the local branches and headquarters of these two organizations throughout the sample years (I coded all the meetings held by other organizations reported in these periodicals as well to have a clear picture of the universe of existing meetings). This project's linguist analyzed these advertisements—which are first written in ancient variations of Dutch and English and

subsequently mutate into proto-Afrikaans and Afrikaans with the passing of the years—and created a database that includes the specific locations, topics planned for discussion, and a description of the key participants in each meeting.[10]

Table 4 summarizes the means and standard deviations for each grassroots organizations meetings when aggregated by district, the level of analysis. The yearly average district number of *Afrikaner Bond* and Farmers Protection Association between 1870 and 1910 is 0.02 and other agricultural organizations is 0.03. The yearly average for all grassroots organizations meetings is 0.28. This is consistent with the historical context: while some districts do not have grassroots organizations meetings at all, some locations such as Paarl and Stellenbosch—two traditional white settler settlements—have during the sample period a total of 133 and 112 meetings of all types of organizations, respectively. The year with most overall meetings is 43.

4.b.4 House of Assembly Members (Cape of Good Hope Colony)

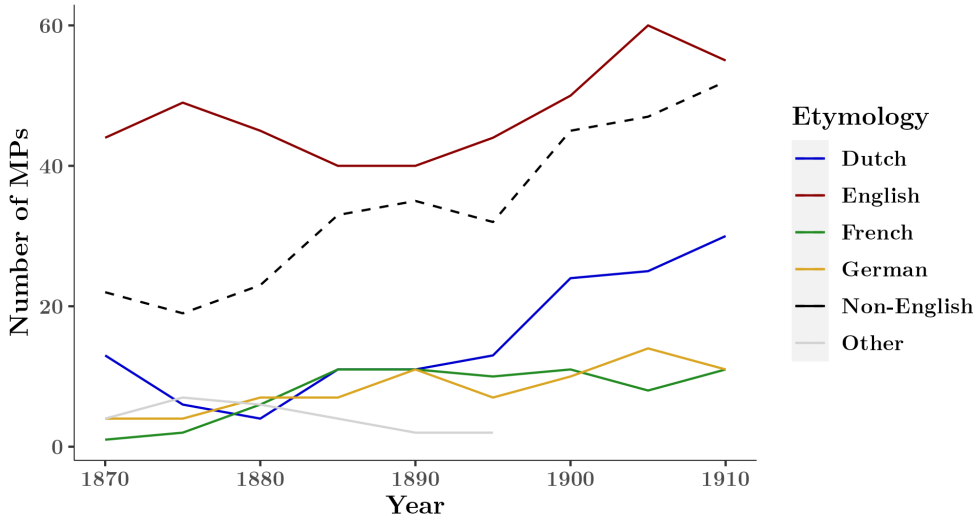
I obtained the list of House of Assembly members' (MPs) in the Cape of Good Hope in *Blue Books* (1870-1885) and *Civil Service Lists* (1886-1910) and background information—such as birth location and occupation—in the *Cape Almanac Series* (1870-1910) and geneological websites. Exceptionally, elected MPs were closely monitored by their local constituents: carpet baggers were unusual, parish pump politics was the rule, and members were known to fear retaliation from their constituents—sometimes even absenting themselves from partaking in controversial votes to avoid animosity. One MP complained about this general situation by

[10]I want to thank Chané Venter for her exceptional contribution in constructing these variables.

stating "[...] that even a parliamentary leader has a right to some opinion [...]" (McCracken, 1967, p. 51).

Further, contemporary reports highlight that supporters of racial separateness policies were prone to vote for candidates with Dutch last names, while opposers usually voted English. Opinions on racial divisions were therefore reflected in the cultural profile of the elected representatives themselves (Davenport, 1960, p. 232). Against this background, MP last names appropriately reflect local voters' own backgrounds, preferences, and interest in large-scale national politics. Figure 5 shows the number of MPs with Dutch and overall non-English last names. The closing gap between English and non-English elected MPs suggests that non-English white populations were increasingly interested in obtaining state representation during the sample years.

Figure 5: MPs Last Name Etymologies



The figure shows the total number of House of Assembly members with last names of any given etymology per year. The closing gap between English and non-English elected MPs suggests that non-English white populations were increasingly interested in obtaining state representation during the sample years.

Table 4 summarizes the share and standard deviations of MPs with different last name etymologies when aggregated by district, the level of analysis. One district in one given year may share a MP with another district, which the database account for. The yearly average district share of MPs with Dutch last names between 1870 and 1910 is 0.18, with English last names is 0.59 and with non-English last names (including Dutch, as well as others) is and other agricultural organizations is 0.41. This shows that English state representation in the Cape of Good Hope is largely predominant throughout the sample years.

4.c Mechanisms: Wages, Agricultural Production, and Land Grants

To analyze how market access changes white settlers' cognitive cost and status I collected three different sets of variables. First, in order to calculate the racial composition of the sample locations and changes in agricultural production, I collected the censuses from the four polities that existed then: the Cape of Good Hope, the Colony of Natal, the Orange Free State, and the Transvaal in different archival repositories across South Africa. I then transcribed the relevant information and merged it into the georeferenced spatial object. The construction of variables required standardizing across languages, changing political units (districts and subdistricts), units of measurement, and racial categories (one census allocates people into 160 different categories).

I obtained racial wage information form the *Blue Books* (1870-1885) and *Statistical Registers* (1886-1909) of the Cape of Good Hope Colony and the *Blue Books* (1870-1893) and *Statistical Year Book* (1894-1909) of the Colony of Natal. These sources separate wages for white and non-white workers in jobs such as farm overseers, farm servants, day laborers,

domestic servants, tradesmen, wagon-riders, among others. The data is at the district level and is available for different sets of years depending on the region, since the data format significantly changes throughout these four decades. For this sample I use two years: I average the data between 1869-1874 and 1886-1891 to account for missing data. I don't use information after 1891, as by then most jobs have been desegregated.

Figure B3 depicts a map of the racial wage gaps by set of occupations for the baseline years 1869-1874, where darker shades represent higher gaps. Table 4 summarizes the wage gap and standard deviations for different types of jobs when aggregated by district, the level of analysis. The yearly average district wage gap for farm jobs is 0.47, for domestic jobs is 0.4, and for all types of jobs is 0.41. Finally, I calculate wheat and maize production by district using the above-mentioned censuses. Table 4 also summarizes the yearly average production of wheat and maize per district, which are 73,894,122 and 11,946,978 liters respectively. Noticeably, the white-dominated production of wheat is higher than the non-white dominated production of maize.

Table 4: Summary Statistics of Outcome Variables

Variable	Observations	Mean	SD	Geo-Unit	Frequency	Polity
Children's First Names Etymology						
Log share Afrikaans	14083	-2.84	0.50	Sub-district	Yearly	All South Africa
Log share English	14083	-2.28	1.02	Sub-district	Yearly	All South Africa
Log share Dutch	14083	-0.12	0.50	Sub-district	Yearly	All South Africa
Newspaper Language						
Log no. Afrikaans	7749	-2.58	1.07	District	Yearly	All South Africa
Log no. English	7749	-2.48	1.20	District	Yearly	All South Africa
Log no. African	7749	-2.93	0.45	District	Yearly	All South Africa
Log no. Indian	7749	-2.99	0.19	District	Yearly	All South Africa
Grassroots Organizations Meetings						
Log no. Afrikaner Bond	4387	-2.95	0.40	District	Yearly	Cape Colony
Log no. Farmers Protection Assoc.	4387	-2.96	0.35	District	Yearly	Cape Colony
Log no. Other Agricultural Org.	4387	-2.94	0.43	District	Yearly	Cape Colony
Log no. All Grassroots Orgs.	4387	-2.81	0.78	District	Yearly	Cape Colony
MP Last Names						
Log share Dutch	809	-2.04	1.28	District	Five-yearly	Cape Colony
Log share English	809	-0.70	1.11	District	Five-yearly	Cape Colony
Log share non-English	809	-1.23	1.34	District	Five-yearly	Cape Colony
Racial Wage Gaps						
Log WG Farm Jobs	115	-0.88	0.63	District	Two Years	Cape Colony & Natal
Log WG Domestic Jobs	105	-0.98	0.40	District	Two Years	Cape Colony & Natal
Log WG All Jobs	118	-0.97	0.49	District	Two Years	Cape Colony & Natal
Agricultural Production (Liters)						
Log Wheat	348	6.62	2.24	District	Four Years	Cape Colony
Log Maize	348	3.30	4.27	District	Four Years	Cape Colony

This table shows summary statistics for main outcome variables between 1870 and 1910. Each set of variables is at either the district or district level, and the time frequency changes depending on the nature of the variable.

5 Empirical Strategy

The goal is to estimate the causal impact of local growth of market access. My main empirical specifications are difference-in-differences estimations that leverage time and geographical variation in local market access increases:

$$\ln Y_{it} = \beta \ln MA(\mathrm{t})_{i,t-1} + \theta X_i' * \gamma_{t-1} + \gamma_i + \gamma_{t-1} + \varepsilon_{it-1} \tag{1}$$

Where $\ln Y_{it}$ is the identity outcome in sub-district or district i and time t. $\ln MA(\mathrm{t})_{i,t-1}$ is total market access in time $t-1$ (when analyzing mechanism I separate this variable by racial profile, $\ln MA(\mathrm{w})_{i,t-1}$ and $\ln MA(\mathrm{nw})_{i,t-1}$, and subsequently I separate the white populations by belief system, $\ln MA(\mathrm{w}_{Protestant})_{i,t-1}$ and $\ln MA(\mathrm{w}_{DRC})_{i,t-1}$), γ_i are location fixed effects, γ_{t-1} time fixed effects, X_i are baseline invariant controls interacted with year dummies,[1] and ε_{it-1} is an error term clustered at the location level to account for serial correlation in $\ln Y_{it}$.[2] The location fixed effects control for all local time invariant characteristics that might be correlated with local market access increases, while the time fixed effects control for general time-variant shocks and policies.

The identifying assumption is that in the absence of local increases in market access, identity choice changes would have followed the same trend everywhere after partialling-out location and time fixed effects. I perform different tests to assess the plausibility of this

[1] I include baseline location total population, farm size, and wheat and maize soil suitability, which are statistically different to zero in locations above/below the market access median (buckwheat, cassava, and other suitabilities are equal to zero). I am working to add other relevant controls and the corresponding balance plots.

[2] The right hand side variables are lagged since construction is reported until the end of the calendar year.

assumption. First, I show that the estimated impacts from changes in market access are robust to controlling for local railroad construction, as defined by different distance buffers around a given location. The fact that estimated impacts of market access are identified from distant changes in the infrastructure network allows to alleviate concerns that local infrastructure placement decisions—which might be correlated with local identity trends—are solely driving the effects.

Second, I create "placebo" railroad lines built after 1910 to reach gold and diamond fields that had not yet been discovered during the sample years. The assumption is that these placebos would have been built as early as possible had the mines' existence been known.[3] If railroad line placement decisions are driven by unobserved determinants of identity choice, proximity to these counterfactual lines would show spurious effects. Conversely, satistically zero coefficients would suggest that the infrastructure network built to reach the mines is plausibly orthogonal to identity choice—suggesting that both the location of mines and subsequent railroad placement decisions are as-good-as-random to contemporary observers. Placebos in the estimations show no significant effects.

5.a Discussion on Further Potential Identification Challenges (and Some Solutions for the Near Future)

Some identification challenges however remain. First, the data does not currently allow me to address the challenge of contemporaneous shocks that may affect local identity trends—such

[3]Chartered companies financed the network to connect the diamond and gold mines to the coasts (and the world) to reap benefits from minerals and ore (Davenport, 1960). Examples are the Natal Railway Company, the Netherlands-South African Railway Company, or Cecil Rhodes' British South Africa Company. It was standard practice for European states to grant special privileges to such commercial organizations to gain control over trade and resources–and secure their part in the Scramble for Africa.

as self-selection to the treatment through migration.[4] Fortunately for this project's purposes however, the historical literature extensively reports that only a handful of MPs were not long-standing locals (McCracken, 1967), that grassroots organizations meetings were most exclusively organized by local communities' long-established settlers (Davenport, 1960), and that Dutch Reformed Church members were the lowest-mobility populations in South Africa, as they were highly tied to the land (Gilliomee, 2009). Only newspapers may be significantly published by migrants.

Second, the location, size, and racial composition of the other markets any given location is being connected to is not random, making them for example recipient of idiosyncratic shocks that may bias the estimates. One solution would be to create a series of counterfactual market access variables, and add them as controls (Boruskyak & Hull, 2022). However, as I unearthed the hundreds of surveys detailing potential counterfactual railroad and road construction plans, I did not find any systematic routes other than the ones that were eventually built. This makes sense in the historical context. In 1870 plans did not extend beyond the Cape Town hinterland. Before the discovery of mines, infrastructure networks were not deemed worth their cost.

If non-random characteristics of other markets are still concerning, a though experiment going through the direction of potential biases may alleviate worries. Suppose some locations are only connected to non-white populations $\ln MA(\mathrm{w})_{i,t-1}$ which may be, for example, more densely populated, poorer, and more vulnerable to rinderpest.[5] The $\ln MA(\mathrm{nw})_{i,t-1}$ treatment

[4] In subsequent versions of this project I will be able to calculate the number of MPs born locally. However, even if MPs are local, maybe the voters that elect them are migrating. Analysis of voter rolls will allow me to verify this in the future.

[5] Due to less access to agriculture investment resources, for example.

would therefore measure market access to locations with poverty associated shocks. Such shocks would however not fuel Afrikaner identity as predicted, since white settlers' social utility would not be threatened by others' adversity. High marginal costs of deviating from the British would, on the contrary, induce British identification. A $\ln MA(\mathrm{w})_{i,t-1}$ wealth shock would produce the opposite effect.

Finally, the data will allow me to further test for pre-treatment parallel trends in potential outcomes in subsequent versions of this project—using twenty or ten years of pre-treatment data, depending on the outcome. The existence of pre-treatment parallel trends would suggest that locations with and without treatment (i.e. above/below median market access) would have followed parallel trends for more periods had the treatment not occurred—dispelling concerns that it is correlated with the outcome trends. I will also include estimations that exclude ports and mines locations—the main objective of the infrastructure expansion—and test whether results change when only including locations that are unintended targets of the infrastructure network expansion.

6 Results

6.a Estimated Impact of Market Access on National Identity

This section reports the average impact of total market access on the different measures for white settlers' national identity choice. Results in Table 5 show main results for the the impact of a one standard deviation greater increase in total market access on white settlers' national identity choice. Results show the activation of a cultural boundary between Afrikaner and British national identities. The main results are robust to controlling for infrastructure placement and placebo infrastructure networks (Tables 6, **??**, and 7). Heterogeneities accounting by number of baseline social groups—white Protestants, white Dutch Reformed Church, and non-white populations—do not show systematic patterns across specifications for identity choice (Table **??**).

6.a.1 Estimated Impact of Market Access on Children's First Names

Panel 5a in Table 5 shows the average effect of aggregate total market access on Afrikaans, English and Dutch children's first names using equation 1. Results show that the impact of a one standard deviation greater increase in total market access increases parent's choice of Afrikaans children's names by 10%, English names by 5%, and decreased Dutch first names by 6%. These results are consistent with the predictions of the social identity model. The largest share of white settlers are predicted to respond to market integration by identifying with the Afrikaner. The rest will identify with the British. Noticeably, first names in Dutch—the most prevalent "influence language" among white settlers before integration—diminishes.

This is consistent with both contemporary accounts of perceptions of Dutch culture in South Africa as well as national narratives typically observed elsewhere during this "age of nationalism" (Hobsbawm, 1990). A central element of Afrikaner national *mythos* was the open repudiation of all things European, as the vision of a real white "African" nation with its own culture, territory, and history was to be separated from any influence perceived to be "imperialistic" (Davenport, 1960; Moodie, 1975; Maré, 1992). On the other hand, settlers that identified with the British would think of Dutch culture as backward and unsophisticated (Davenport, 1960). Observing such trends permeate children's names highlights the broad reach of nationalism well into people's private life.

6.a.2 Estimated Impact of Market Access on Newspaper Language

Panel 5b in Table 5 reports the average effect of aggregate total market access on the total number of publications in Afrikaans, English, different African, and various Indian languages. The impact of a one standard deviation greater increase in district market access on Afrikaans publications is 24% and on English publications is 23%. Both estimates are high and significant. On the other hand, the effect on African and Indian languages is not significant. These results suggest that integration of markets is making people converge around the two strongest national identities that are developing, strengthening the boundary between supporters of the nascent Afrikaner nation on the one hand and English on the other (Deutsch, 1953).

This variable comes with a caveat. The language of newspapers reflects the communication demands of the general population, not only white settlers; I therefore cannot guarantee that

the increase in Afrikaans or English publications reflects white settlers national identity choice. Results do however allow to track how the population is separating into two nations as market integration increases, Afrikaans and English becoming the leading communication spaces. Further, results also highlight how the two languages that consitute the backbone of white national groups largely dominate the social space, relegating the various African and Indian languages spoken by the numerically superior non-white populations of South Africa to a secondary level.

6.a.3 Estimated Impact of Market Access on Grassroots Organizations (Cape)

Panel 5c in Table 5 shows the average effect of aggregate market access on the number of meetings of the grassroots organization Afrikaner Bond (AB), the Farmer Protection Association (*Boeren Beschermings Vereeniging* or "BB") and other non-political agricultural organizations, using equation 1. The impact of a one standard deviation greater increase in district market access on Farmers Protection Associations and other agricultural organizations is high and significant, respectively 32% and 18.4%. The effect is not significant and negative for the meetings of the political organization Afrikaner Bond (AB). The organizations that primarily focus on the economic aspects of British institutions are more responsive to increases in total market access than the Afrikaner political one.

This contrast is compatible with the predictions of the social identity model. First, the Afrikaner Bond (AB) captures a more radical nationalist Afrikaner response when compared to parents choice of children's names or the language of newspaper publications analyzed above. As a militant organization, one could expect it to be especially triggered by contact

and mobility of non-white populations brought about by the integration of markets—not necessarily the total aggregate market access treatment. Second, this sample is restricted to the British-run Cape colony, where white settlers are more prone to develop pro-British sentiments; which is perhaps what also explains why the Afrikaner counterpart is more radical (Davenport, 1960).

These considerations beg the question about how this measure of national identity responds to increases in market access when separated by white and non-white populations, which I analyze below. Finally, Table B1 in the appendix further analyzes the topics of all the meetings advertised: a one standard deviation increase in market access heightens advertisements that discuss economics by 30% and the British in 11%.

6.a.4 Estimated Impact of Market Access on MPs Last Names (Cape)

Panel 5d in Table 5 shows the estimates of the effect of total market access on the House of Assembly Members' last name etymology, using equation 1. Increases in aggregate market access has no significant effect on the election of MPs with Dutch or English last names; although Dutch last names seem to increase and English to decrease. Increases in market access does however have a large and significant effect on the election of MPs with non-English last names. A one standard deviation increase in market access heightens the election MPs with non-English last names by 30%, where MPs with French last names lead this effect. These results suggest that the integration of markets is decreasing relative English dominance at the state level, allowing new actors into positions of power.

Table B2 in the appendix further explores the occupational profile of the elected House

of Assembly members. There is a significant and large increase in the number of MPs that are farmers: a one standard deviation increase in market access increases farmer MPs in 33%. The effect on military MPs is significant but negative, the estimate shows a 29% decrease. MPs that are professionals, businessmen, and lawyers seem to decrease with increases in market access; these estimates are however not significant. These trends are congruent with the social identity model predictions in the South African colonial context. Market integration galvanizes the formation of a nation that is tied to the land and pushes for an official position in the state.

Table 5: Estimated Impact of Market Access on National Identity Choice

(a) Dependent Variables: Children's Names

	Afrikaans (1)	English (2)	Dutch (3)
Market Access	0.096***	0.051*	-0.063**
	(0.033)	(0.028)	(0.026)
Observations	14,083	14,083	14,083
R^2	0.190	0.301	0.159
Sub-District FE (687)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans (1)	English (2)	African (3)	Indian (4)
Market Acccess	0.233**	0.228***	0.167	0.040
	(0.092)	(0.073)	(0.128)	(0.036)
Observations	7,749	7,749	7,749	7,749
R^2	0.726	0.766	0.482	0.168
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

(c) Dependent Variables: Grassroots Orgs. Meetings

	Afrikaner B. (1)	Farmer Assoc. (2)	Other (3)
Market Acccess	-0.103	0.315**	0.184*
	(0.123)	(0.129)	(0.100)
Observations	4,387	4,387	4,387
R^2	0.120	0.217	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

(d) Dependent Variables: MPs Last Names

	Dutch (1)	English (2)	non-English (3)
Market Access	0.176	-0.204	0.300***
	(0.169)	(0.141)	(0.100)
Observations	809	809	809
R^2	0.368	0.476	0.526
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. Panel 5a displays results for the share of children's first names of a given etymology. Panel 5b displays results for the number of newspaper publications in different languages. Panel 5c displays results for the number of meetings for the pro-Afrikaner *Afrikaner Bond*, the pro-British *Farmers Protection Association*, and other agricultural organizations. Panel 5d displays results for the share of elected MPs' last names of a given etymology. Panel 5a is at the sub-district level. Panel 5b-5d are at the district level. Panels 5a-5c are yearly panels. Panel 5d is a five-yearly panel. All variables are in logarithms. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

6.a.5 Heterogeneities by Initial Populations: White Protestant, White DRC, and Non-White Populations

Heterogeneous effects by initial populations do not show systematic trends across the national identity variables; the significant coefficients however do show some predictable trends. Table **??** shows these results. Only Panels B3a and Panel B4b show significant heterogeneities when interacting by initial number of white Dutch Reformed Church and Protestant members—most likely descendants from the first—Dutch East India Company (VOC)—and second—British—migration waves. Increases in market access in locations with high initial numbers of white Protestants decreases Afrikaans children's first names and the election of Dutch and non-English MPs, and increases English MPs. Heterogeneity by initial white DRC populations increases number of Dutch MPs.

Interestingly, increases in market access heterogeneities with respect to number of initial non-white populations shows a significant and positive estimate for newspapers in Afrikaans and Afrikaner Bond meetings; and a significant and negative estimate for election of MPs with English last names.

6.b Addressing Identification Challenges

6.b.1 Controlling for Infrastructure Placement

Tables 6a, 6b, 7c, and 7d show in column (1) the baseline result reported above, and in columns (2) to (4) estimations that control for the existence of a road or railroad built within 50km, 100km, 200km of each location (either district or sub-district) in every year. The estimates

that control for changes in local infrastructure placement to specifically capture the effects of non-local market access variation, as defined by the different distance buffers. This effect at times slightly increases at 50km and then smoothly diminishes as the railroads or roads are further away from the location. Both significance and magnitude diminish, until becoming smaller or insignificant at the 200km buffers. This means that there is a market access effect that is exogenous to local infrastructure placement and that spillovers are strong even when there is no infrastructure construction nearby.

Table 6: Estimated Impact of Market Access on National Identity Choice Controlling for Local Infrastructure Placement

(a) Dependent Variables: Children's Names

	Afrikaans				English				Dutch			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Market Access	0.096***	0.111***	0.092**	0.065*	0.051*	0.057*	0.044	0.035	-0.063**	-0.048*	-0.041	-0.054*
	(0.033)	(0.038)	(0.037)	(0.037)	(0.028)	(0.034)	(0.032)	(0.030)	(0.026)	(0.028)	(0.029)	(0.028)
Observations	14,083	14,083	14,083	14,083	14,083	14,083	14,083	14,083	14,083	14,083	14,083	14,083
R^2	0.190	0.190	0.190	0.190	0.301	0.301	0.301	0.301	0.159	0.159	0.159	0.159
Sub-District FE (687)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans				English			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
MA (t)	0.233**	0.256**	0.222**	0.224**	0.228***	0.219***	0.224***	0.197***
	(0.092)	(0.106)	(0.101)	(0.094)	(0.073)	(0.082)	(0.076)	(0.073)
# District	189	188	188	188	189	188	188	188
Observations	7,749	7,708	7,708	7,708	7,749	7,708	7,708	7,708
R^2	0.726	0.726	0.726	0.726	0.766	0.758	0.758	0.758
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓	✓	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1, controlling by local infrastructure placement. Panel 6a displays results for the share of children's first names of a given etymology. Panel 6b displays results for the number of newspaper publications in different languages. Columns (1) reproduce the baseline results for each of the different significant dependent variables. Columns (2) to (4) show estimations that control for the existence of a road or railroad built within 50km, 100km, 200km of each location (sub-district or district) in every given year. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table 6 *(continued)*: Estimated Impact of Market Access on National Identity Choice
Controlling for Local Infrastructure Placement

(c) Dependent Variables: Grassroots Organizations Meetings

	Farmer's Protection Assoc. (BB)				Other Agricultural Orgs.			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Market Access	0.315**	0.304**	0.296**	0.297**	0.184*	0.171*	0.173*	0.179*
	(0.129)	(0.128)	(0.123)	(0.123)	(0.100)	(0.098)	(0.096)	(0.098)
Observations	4,387	4,387	4,387	4,387	4,387	4,387	4,387	4,387
R^2	0.217	0.218	0.218	0.218	0.388	0.388	0.388	0.388
District FE (107)	✓	✓	✓	✓	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓	✓	✓	✓	✓

(d) Dependent Variables: MPs Last Names

	non-English			
	(1)	(2)	(3)	(4)
MA (total)	0.300***	0.314***	0.335***	0.303***
	(0.100)	(0.109)	(0.105)	(0.100)
Observations	809	809	809	809
R^2	0.526	0.526	0.527	0.526
District FE (106)	✓	✓	✓	✓
Year FE (9)	✓	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1, controlling by local infrastructure placement. Panel 7c displays results for the number of grassroots organizations meetings. Panel 7d displays results for the share of elected MPs' last names of a given etymology. Columns (1) reproduce the baseline results for each of the different significant dependent variables. Columns (2) to (4) show estimations that control for the existence of a road or railroad built within 50km, 100km, 200km of each location (sub-district or district) in every given year. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

6.b.2 Placebo Infrastructure

Tables 7a, 7b, 7c, and 7d include a "placebo" infrastructure network that varies yearly. I construct this variable by creating an iinfrastructure network of lines that were built *after* the end of the sample years to reach mines that had yet not been discovered. The assumption is that these lines would have been built as early as possible had the mines' existence been known. Consistently throughout the Panels, estimates are insignificant; the exception is a s significant and negative placebo coefficient impacting the election of MPs with English last names. The overall trends throughout the Panels however highlight that line placement decisions are likely not driven by unobservable determinants of identity; as they do not have spurious impacts on national identity choice.

Table 7: Estimated Impact of Market Access on National Identity Choice Controlling for Placebo Infrastructure

(a) Dependent Variables: Children's Names

	Afrikaans (1)	English (2)	Dutch (3)
MA (total)	0.095***	0.052*	-0.063**
	(0.033)	(0.028)	(0.026)
Placebo	0.056	-0.062	0.029
	(0.050)	(0.063)	(0.087)
Observations	14,083	14,083	14,083
R^2	0.190	0.301	0.159
Sub-District FE (687)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans (1)	English (2)	African (3)	Indian (4)
MA (t)	0.229**	0.226***	0.173	0.036
	(0.093)	(0.073)	(0.131)	(0.037)
Placebo	0.109	0.072	-0.076	-0.024
	(0.125)	(0.130)	(0.080)	(0.024)
Observations	7,708	7,708	7,708	7,708
R^2	0.726	0.758	0.489	0.103
District FE (188)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

(c) Dependent Variables: Grassroots Orgs. Meetings

	AB (1)	BB (2)	Other (3)
MA (total)	-0.178	0.283**	0.258
	(0.138)	(0.124)	(0.163)
Placebo	0.111	0.048	-0.111
	(0.079)	(0.076)	(0.142)
Observations	4,387	4,387	4,387
R^2	0.120	0.218	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

(d) Dependent Variables: MPs Last Names

	Dutch (1)	English (2)	non-English (3)
MA (total)	0.181	-0.213	0.300***
	(0.168)	(0.136)	(0.100)
Placebo	0.341	-0.572***	0.035
	(0.216)	(0.205)	(0.198)
Observations	809	809	809
R^2	0.371	0.484	0.526
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1, controlling for placebo infrastructure. Each Panel contains a different set of national identity choice variables, following the same structure as the main results in Table 5. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

7 Mechanisms

This section explores the mechanisms that drive white settlers' choice of national identity amid eroding traditional social structures. Following the social identity model in Section 3, I analyze the impact of market access on white settlers' social utility which I hypothesize is a function of two main components. The first is the distance to the prototypical attributes of other social category members, such as "white-farmers", changes in the salience of which I capture by measuring market access to populations with different racial profiles and belief systems. The second is the relative material status of different social categories, such as "white-" relative to "non-white farmers", variation in which I measure using changes in racial wage gaps and agricultural production by racial group.

Results show that increases in market access to non-white populations induces white settlers to identify more with the Afrikaner and less with the British. Increases in market access to white populations has the opposite effect. However, heightened belief system salience has either no or counterintuitive effects: increases in market access to other white Dutch Reformed Church members—white settlers' baseline religion—decreases Afrikaner and increases British identity choice. While white Protestant salience does decrease British identification, highlighting that "foreign" ideologies are not particularly liked, it does not impact Afrikaner identity choice. National identity choice thus responds most exclusively to racial not belief system salience. National boundary activation is driven by race.

Results further show that changes in racial salience coincide or even galvanize changes in racial relative status, particularly the "white farmer" social category—which most white

settlers identify with—relative to the "non-white farmer". Higher increases in market access to non-white populations decrease the racial wage gap, reducing racial relative status and increasing the marginal value of institutionalizing non-white populations' repression, triggering Afrikaner identification (market access' impact on increases in non-white agricultural production generate a similar effect). Increases in market access to white populations increase the racial wage gap, heightening racial relative status and the marginal cost of not supporting British nation, triggering British identification.

In sum, results show that the impact of market access on the extent of local racial mobility—or desegregation—shapes individuals' choice of large-scale national norms and institutions.

7.a Estimated Impact of Market Access to Different Population Profiles: Race and Belief Systems (Religion)

In this section I analyze the effect of increases in market access to populations with different racial profiles and belief systems on the various national identity variables. I create two different sets of market access measures. First, a market access variable that calculates the costs of trade to baseline white ($N_j(w)$) and non-white ($N_j(nw)$) populations (see section 4.a.3 for details). I then further split initial white populations into Dutch Reformed Church (DRC) ($N_j(w_{DRC})$) and other Protestants ($N_j(w_{Protestant})$) populations, in order to characterize white populations that are most likely descendants from the first—Dutch East India Company (VOC)—and second—British—migration waves which arguably possess differing

belief systems mediated by their religion.[1]

I test whether white settlers' national identity choice is driven by changes in the salience of people with different prototypical attributes in dimensions such as race, belief systems, both, or neither. Although the rise of Afrikaner nationalism is frequently analyzed as driven by the rejection of the incoming British based on religious or linguistic differences (Davenport, 1960; Bundy, 1979; Gilliomee, 2009), the data disproves this hypothesis. Results show a more nuanced story: the local salience of white Protestants does not explain the Afrikaner rise; racial salience does. Moreover, increases in market access to other DRC members decreases Afrikaner identity. This supports the hypothesis that Afrikaner nationalism is fueled by perceived threat of desegregation, not newfound community.

7.a.1 Estimated Impact of Market Access to Different Racial Profiles

Table 8 shows results for the average effect of aggregate market access to white and non-white populations on the different national identity variables. Panels 8b,8c, and 8d show similar patterns. A one standard deviation greater increase in market access to white populations decreases white settlers' Afrikaner national identity choice and increases English identity choice. Afrikaner Bond meetings and the election of MPs with Dutch last names decrease by 100% and 109%, respectively. Newspaper publications in English, British-oriented Farmers Protection Association meetings, and MPs with English last names increase in 44%, 90%, and 65%, respectively. The salience of white populations therefore prompts white settlers to identify with the English nation.

[1] I use religion as a measure to identify these two groups, since white or European population are not distinguishable by racial or cultural attributes in any of the censuses during the sample years—and beyond.

A one standard deviation greater increase in market access to non-white populations shows the opposite effect. The treatment increases Afrikaner and decreases English national identity choice. Newspaper publications in Afrikaans, Afrikaner Bond meetings, and the election of MPs with Dutch last names increase in 37%, 84%, 130%, respectively. British-oriented Farmers Protection Association meetings and the election of MPs with English last names decrease in 55% and 65%, respectively. The salience of non-white populations prompts white settlers to identify with the Afrikaner nation. Unlike the main results (Panels 5c and 5d of Table 5), treatment effects on Afrikaner Bond meetings and election of MPs with Dutch last names are high and significant.

Results in Table 8 are consistent with the social identity model predictions in Section 3. The local salience of populations with different racial profiles is driving the activation of a within-white cultural boundary. Increases in market access to white populations tilts white settlers' identification towards the English rather than the Afrikaner, suggesting the cognitive costs of joining the former are lower than those associated with the latter. This hints to the fact that increased salience of outside white populations is not enough to prompt white settlers to support the nascent Afrikaner. The Afrikaner segregationist agenda is redundant if the prototypical racial attributes of society remain the same, or predominantly white. Identifying with British norms and institutions is less costly.

The increase in salience of non-white populations highlights a completely different story. White settlers identify with the Afrikaner rather than English nation, as the cognitive costs of joining the latter are now higher than the former. British norms and institutions would allow for the prototypical racial attributes of people white settlers share social categories

and roles with to freely change—such as "white farmer". Supporting the Afrikaner agenda of "separateness" is the white settlers' strategy to decrease this cognitive cost. However, the fact that the effects on Afrikaner Bond meetings and the election of MPs with Dutch last names are only significant in Panels 8c and 8d (and not in Panels 5c and 5d) suggests that joining the Afrikaner is not taken lightly; it is inherently costly to oppose the British.

Finally, results in Panel 8a do not show these same patterns as the other national identity variables in 8b,8c, and 8d; white settler choice of children first names is not responding to increases in salience of populations with different racial profiles. This sample is most representative of the Boer Republics, which may indicate that the salience of racial prototypical attributes is not the most relevant for identity formation at the very personal level: parent's choice of children's names.

Table 8: Estimated Impact of Market Access on National Identity Choice Separating by Population Racial Profiles

(a) Dependent Variables: Children's Names

	Afrikaans (1)	English (2)	Dutch (3)
MA (white)	0.064	-0.053	-0.012
	(0.130)	(0.115)	(0.122)
MA (non-white)	0.035	0.105	-0.051
	(0.126)	(0.114)	(0.121)
Observations	14,083	14,083	14,083
R^2	0.190	0.301	0.159
Sub-District FE (687)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans (1)	English (2)	African (3)	Indian (4)
MA (white)	-0.138	0.449***	0.332	-0.004
	(0.201)	(0.169)	(0.271)	(0.068)
MA (non-white)	0.365**	-0.180	-0.121	0.038
	(0.182)	(0.145)	(0.231)	(0.086)
Observations	7,749	7,749	7,749	7,749
R^2	0.726	0.767	0.484	0.168
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

(c) Dependent Variables: Grassroots Orgs. Meetings

	Afrikaner B. (1)	Farmer Assoc. (2)	Other (3)
MA (white)	-0.994***	0.902***	0.111
	(0.306)	(0.286)	(0.320)
MA (non-white)	0.839***	-0.549***	0.073
	(0.255)	(0.170)	(0.295)
Observations	4,387	4,387	4,387
R^2	0.124	0.220	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

(d) Dependent Variables: MPs Last Names

	Dutch (1)	English (2)	non-English (3)
MA (white)	-1.09***	0.652**	-0.343
	(0.289)	(0.281)	(0.303)
MA (non-white)	1.31***	-0.871***	0.651*
	(0.326)	(0.287)	(0.329)
Observations	809	809	809
R^2	0.384	0.482	0.528
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in market access to populations of different racial profiles on each set of dependent variables using equation 1. Each Panel contains a different set of national identity choice variables, following the same structure as the main results in Table 5. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

7.a.2 Estimated Impact of Market Access to Belief Systems and Racial Profiles

Table 9 shows results for the average effect of aggregate market access separating white populations into Dutch Reformed Church (DRC) and other Protestant members, and non-white populations on the different national identity variables. A one standard deviation greater increase in market access to non-white populations shows the same patterns as above for all the panels. Market access to white populations with different religious backgrounds reveal interesting results. A one standard deviation greater increase in market access to white DRC populations is systematically either not significant or significantly decreases Afrikaner national support. Panels 9c and 9d show that Afrikaner Bond meetings and the election of MPs with Dutch last names decrease in 61% and 65%, respectively.

This effect may seem paradoxical. Since white settlers are most likely to be DRC members, it might seem reasonable to expect that Afrikaner national identity would grow as market access to similar ideologically minded peers grows. Negative effects however support the hypothesis that increases in market access to white populations do not threaten white settlers enough to galvanize their support for the Afrikaners. On the contrary, a one standard deviation greater increase in market access to initial white DRC populations heightens white settler choice of English national identity. Panels 9c and 9d show that British-oriented Farmer Protections Association meetings and election of MPs with English last names increases in 102% and 89%.

Panels 9c, and 9d do however show that a one standard deviation greater increase in market access to white Protestant populations decreases English identity choice. British-

oriented Farmer Protections Association meetings and election of MPs with English last names decerease in 108% and 82%, respectively. These results suggest that white settlers may be indeed reacting negatively to influxes of white populations with different belief systems by not supporting British institutionality and representatives. This however does not clearly translate into Afrikaner support. Finally, Panel 9a does not show any significant Afrikaans children first names. This sample is most representative of the Boer Republics, which is why it might differ from the others Panels.

As a last exercise, I interact the different market access variables with the baseline local number of white DRC members. Tables B8, B9 and B10 in the appendix shows that heterogeneous effects by initial DRC members is either not significant or negatively impacts Afrikaner identity choice. Only in Table B11, the heterogeneous effect on market access to protestant populations increases MPs with Dutch last names. Finally, tables B9 and B10 show that heterogeneity may also affect English last names, negatively if market access to protestants or non-whites increase (who are perceived as threats) and positively if market access to other DRC members increases.

Table 9: Estimated Impact of Market Access on National Identity Choice by Religion and Non-White Populations

(a) Dependent Variables: Children's Names

	Afrikaans (1)	English (2)	Dutch (3)
MA (Protestant)	0.165	0.265	-0.016
	(0.181)	(0.180)	(0.208)
MA (DRC)	-0.072	-0.133	-0.023
	(0.184)	(0.176)	(0.206)
MA (non-white)	0.010	-0.070	-0.028
	(0.078)	(0.073)	(0.073)
Observations	14,083	14,083	14,083
R^2	0.190	0.302	0.159
Sub-District FE (687)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans (1)	English (2)	African (3)	Indian (4)
MA (white Protestant)	-0.037	-0.020	-0.343	-0.631
	(0.261)	(0.186)	(0.310)	(0.443)
MA (white DRC)	-0.085	0.305	0.491*	0.438
	(0.265)	(0.217)	(0.249)	(0.342)
MA (non-white)	0.343**	-0.013	0.070	0.223
	(0.151)	(0.082)	(0.201)	(0.158)
Observations	7,749	7,749	7,749	7,749
R^2	0.727	0.767	0.485	0.172
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

(c) Dependent Variables: Grassroots Orgs. Meetings

	Afrikaner B. (1)	Farmer Assoc. (2)	Other (3)
MA (white Protestant)	-0.389	-1.08*	-0.173
	(0.275)	(0.631)	(0.354)
MA (white DRC)	-0.607***	1.02***	0.174
	(0.203)	(0.372)	(0.280)
MA (non-white)	0.745**	0.390	0.185
	(0.324)	(0.398)	(0.451)
Observations	4,387	4,387	4,387
R^2	0.123	0.221	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

(d) Dependent Variables: MPs Last Names

	Dutch (1)	English (2)	non-English (3)
MA (white Protestant)	-0.664	-0.820**	-0.416
	(0.440)	(0.405)	(0.365)
MA (white DRC)	-0.643***	0.887***	-0.094
	(0.242)	(0.233)	(0.276)
MA (non-white)	1.41***	-0.289	0.775**
	(0.362)	(0.320)	(0.312)
Observations	809	809	809
R^2	0.389	0.489	0.529
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in market access to white member of different religions and non-white populations on each set of dependent variables using equation 1. Religious groups correspond to non-English and English populations. Each Panel contains a different set of national identity choice variables, following the same structure as the main results in Table 5. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

7.b Changes in Relative Status

The social identity model highlights a second key driver of national identity choice: changes in social categories' relative status—such as the material payoffs of "non-white farmers" relative to "white farmers". Increases in racial salience—analyzed above—may coincide with changes in racial relative status. First, market access to non-white populations may close the racial wage gap; as the non-white labor pool expands, employers secure cheaper newly available non-white labor while white workers accept lower relative wages to ensure employment. Markets relax segregation. Conversely, if the labor pool remains predominantly white, the racial wage gap may increase as employers must hire white workers that demand higher relative wages. Markets reinforce segregation.[2]

Following the social identity model, these two scenarios may have a differing effect on white settlers' national identity choice (see Section A for a mathematical representation). If market access to non-white populations galvanizes desegregation, white settlers' marginal value of institutionalizing the repression of non-white populations increases, identifying with the Afrikaner since the social cost of exerting repression under British institutions is intolerable. If, on the other hand, market access to white populations relaxes desegregation, white settlers' status remains unchanged and there are no incentives for repression; further, the marginal costs of deviating from British norms and institutions that dominate world markets outweigh the benefits, tilting identification towards the British.

Second, agricultural production's susceptibility to desegregate as market access increases

[2]This framework is closely related to the "split-market" literature that analyzes racially segmented unions (Bonacich, 1972, 1973, 1976; Farley, 1988).

may further explain the rise of Afrikaner nationalism. As both "white-" and "non-white farmers" benefit from deeper integration, increased competition may heighten white farmers' marginal value of supporting institutionalized repression and identify with the Afrikaner—to control and retain non-white labor either from other farms or British-owned mines (Acemoglu & Robinson, 2006; Samuels & Thomson, 2020). This mechanism may be assessed by measuring the impact of increases in market access in traditionally non-white and white agricultural districts on agricultural output. A crowding-out of white production would decrease racial status and trigger Afrikaner identification.

In sum, the differential impact of markets on the desegregation of the economy via variation in racial relative status drives white settlers' national identity choice—whether deepening markets change the racial composition of the labor force impacting racial wage gaps, or whether they benefit farmers of different racial profiles' relative agricultural production. Namely, racial mobitlity shapes white settlers' choice of large-scale institutions, which are proposed and enforced by nations. Results below discuss the empirical evidence in support of these mechanisms.

7.b.1 Estimated Impact of Market Access on Racial Wage Gaps

Table 10 shows the results of the effect of a one standard deviation greater increase in market access on the racial wage gaps in farm, domestic, and all jobs. Panel 10a shows that the estimated impact of market access on racial wage gaps does not have a clear significant effect on wage gaps. This may be the case because the racial wage gap is hypothesized to vary depending on the change in the relative availability of white and non-white workers in the

local labor market. Panel Panel 10b therefore divides market access to white and non-white populations. Consistent with the theory, the impact of market access to white populations increases the wage gaps for farming jobs by 140%. Conversely, the impact of market access to non-white populations decreases these wage gaps by 137%.

These results highlight that increases in market access to populations of different racial profiles is not only varying the local salience of non-white populations, but is changing in divergent ways the relative material payoffs of white settlers as well. Noticeably, increases in market access to non-white and white populations that heighten white settlers' choice of Afrikaner and British identity (see Table 8) are congruent with results in Panel 10b that show that changes in the racial composition of the local labor market also impacts racial relative status. The higher the economic desegregation, the higher the marginal value of repression and the higher Afrikaner identification. The lower the economic desegregation, the higher the cost of deviating from the British, the higher British identification.

7.b.2 Estimated Impact of Market Access on Traditional Agricultural Production by Racial Profile

Table 11 shows results of the effect of a one standard deviation greater increase in total market access on total traditional agricultural production. I focus on two main traditional staples: wheat and maize. The first is a traditional white staple while the second is a traditional non-white one—mainly black African (Bundy, 1979): the baseline census shows that whites produce 89% of total liters of wheat and non-whites produce 80% of total liters of maize. Each estimation includes two interactions that capture traditional agricultural districts by

Table 10: Estimated Impact of Market Access on Racial Wage Gaps by Job Type

(a) Dependent Variables: Racial Wage Gaps

	Farm (1)	Domestic (2)	All (3)
MA (total)	-0.083	0.080	-0.117
	(0.210)	(0.334)	(0.168)
# District	78	74	79
Observations	115	105	118
R^2	0.541	0.714	0.730
District fixed effects	✓	✓	✓
Year fixed effects (4)	✓	✓	✓

(b) Dependent Variables: Racial Wage Gaps

	Farm (1)	Domestic (2)	All (3)
MA (white)	1.49**	1.40	1.09*
	(0.652)	(0.981)	(0.645)
MA (non-white)	-1.37**	-1.72	-0.948
	(0.588)	(1.06)	(0.636)
# District	78	74	79
Observations	115	105	118
R^2	0.912	0.869	0.899
District fixed effects	✓	✓	✓
Year fixed effects (2)	✓	✓	✓

This table displays the estimated impact of market access on the racial wage gap of farm, domestic, and wagon (trade) jobs. This information is taken from four different census years (1865, 1875, 1890, and 1904). Panel 10a reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. Panel 10b reports estimates of the impact of a one standard deviation greater increase in market access to populations of different racial profiles on each set of dependent variables using equation 1. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

race. The first is a variable that is equal to one if the district is above the 75th percentile in non-white production of wheat (maize); the second is equal to one if it is above the 75th

percentile in white production of wheat (maize).

Results show that a one standard deviation greater increase in total market access increases both total wheat and maize production. However, these effects vary by area. Wheat production outside traditional production areas increases the most, by 44%, which suggests that agricultural production expands on the extensive margin as market access increases. Wheat production also increases in non-white wheat producing areas, by 31%, which is evidence of an intensive margin increase as well. Market access however negatively impacts wheat production in traditional white areas by 21%. This suggests that white traditional agriculture is crowded-out by both production in an expanding agricultural frontier and non-white agricultural areas.

Results in column (2) suggest a similar trend for maize. A one standard deviation greater increase in market access heightens maize production in non-white maize traditional production areas by 36%. Results are not significant but maintain the same sign for the agricultural frontier and white production. These results suggest that white settlers lose relative status in the agricultural production sector; markets are bringing about the desegregation of agriculture that both expands the agricultural frontier and provides novel production opportunities to non-whites, both of which deprive white settlers of traditional cheap or coerced non-white labor. Joining the Afrikaner allows them to institutionalize repression of non-white populations, they reject British norms and institutions.

Table 11: Estimated Impact of
Estimated Impact of Market Access on Traditional Agricultural Production

	Wheat (1)	Maize (2)
Market Access	0.441***	0.237
	(0.165)	(0.210)
MA × Traditional non-white wheat	0.312***	
	(0.096)	
MA × Traditional white wheat	-0.218***	
	(0.068)	
MA × Traditional non-white maize		0.356***
		(0.102)
MA × Traditional white maize		-0.031
		(0.066)
Observations	348	348
R^2	0.637	0.611
District fixed effects (107)	✓	✓
Year fixed effects (4)	✓	✓

This table displays the estimated impact of market access on total wheat and maize production. This information is taken from four different census years (1865, 1875, 1890, and 1904). All variables are in logarithms. Interactions capture traditional agricultural districts by race. The first (second) set are variables that are equal to one if the district is above the 75th percentile in non-white and white production of wheat (maize). All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

8 Concluding Remarks

This project explores whether deepening markets make or break nations. I study the impacts of increases in market integration on white settlers' national identity choice in colonial South Africa. For causal identification, I exploit time and geographical variation in increases in local market access, a reduced-form expression derived from general equilibrium trade theory. I estimate the impact of market access on parent's choice of children's first names, the language of newspaper publications, grassroots organizations meetings, and members of parliament's last names. I find that higher increases in market access lead white settlers to choose either British or Afrikaner national identify, effectively activating a cultural boundary between them.

In order to analyze the mechanisms behind this results, I explore market access' differential impacts on racial mobility. Higher racial mobility decreases white settlers' social utility, which is a function of racial relative status and cognitive distance, increasing the marginal value of institutionalizing non-white populations' repression, prompting identification with the Afrikaner—who support this policy. The absence of racial mobility does not decrease white settlers' social utility, increasing the marginal cost of not identifying with the dominant British—who prohibit institutionalized racial segregation. Finally, I find no evidence that other normative disagreements, such as religious beliefs, drive this national divide.

A White Settler's Maximization Problem[1]

The white settler maximizes their social utility by choosing an optimal national identity $n \leq n_B$ which may be further or closer to n_B, the dominant British national identity. The further the national identity n is from n_B, the more Afrikaner. National identity choice reflects white settler's preference of large-scale norms and institutions: identifying with the Afrikaner implies supporting institutionalized segregation—or the repression of non-white populations—while identifying with the British implies opposing it. The white settler's social utility function may be represented as follows:

$$\Pi = y - \underbrace{L(w - r + rf(n_B - n))}_{\text{Wage bill}} - \frac{1}{2} \underbrace{\alpha(n_B - n)^2}_{\substack{\text{Loss function of} \\ \text{deviating from} \\ \text{British}}} \tag{2}$$

Where Π are white settler's total payoffs, y is income, and there is a wage bill where L is a parameter that captures increases in desegregation in the economy, such as higher market access to non-white populations (which may change the labor pool) or increases in market access to non-white agricultural districts (that may increase trade opportunities), w are wages, and r are the gains from institutionalizing the repression of non-white populations. The term $rf(n_B - n)$ is the desutility of using repression to drive down non-white workers' wages (or agricultural production), with $f'(\cdot) < 0$. The cost of repression is higher the closer n is to the British nation n_B, since British norms and institutions do not allow for institutionalized segregation—repression is socially intolerable under this social paradigm.

[1] I want to thank James Robinson for suggesting this model to represent the white settlers' national identity choice.

White settler's payoffs also include a loss function from deviating from British norms and institutions, which dominate world markets. α is a parameter that captures increases in economic opportunities for segregation, such as increases in market access to white populations (that do not change the labor pool) or agricultural benefits tailored to white populations (such as land grants, to be explored in subsequent version of this project).

The white settler's F.O.C. and S.O.C. are as follows:

$$\begin{aligned} \frac{\partial \Pi}{\partial n} &= -Lrf'(\cdot) + \alpha(n_B - n) = 0 \\ \frac{\partial^2 \Pi}{\partial n^2} &= -Lrf''(\cdot) - \alpha \leq 0 \end{aligned} \tag{3}$$

Totally differentiating the F.O.C. with respect to national identity choice, n, and the parameters L and α, results in the following expression:

$$-\mathrm{d}Lrf'(\cdot) - \mathrm{d}n(Lrf''(\cdot) + \alpha) + \mathrm{d}\alpha(n_B - n) = 0 \tag{4}$$

Simplifying further to analyze comparative statics:

$$\begin{aligned} \frac{\mathrm{d}n}{\mathrm{d}L} &= \frac{-rf'(\cdot)}{Lrf'(\cdot) + \alpha} < 0 \\ \frac{\mathrm{d}n}{\mathrm{d}\alpha} &= \frac{n_B - n}{Lrf'(\cdot) + \alpha} > 0 \end{aligned} \tag{5}$$

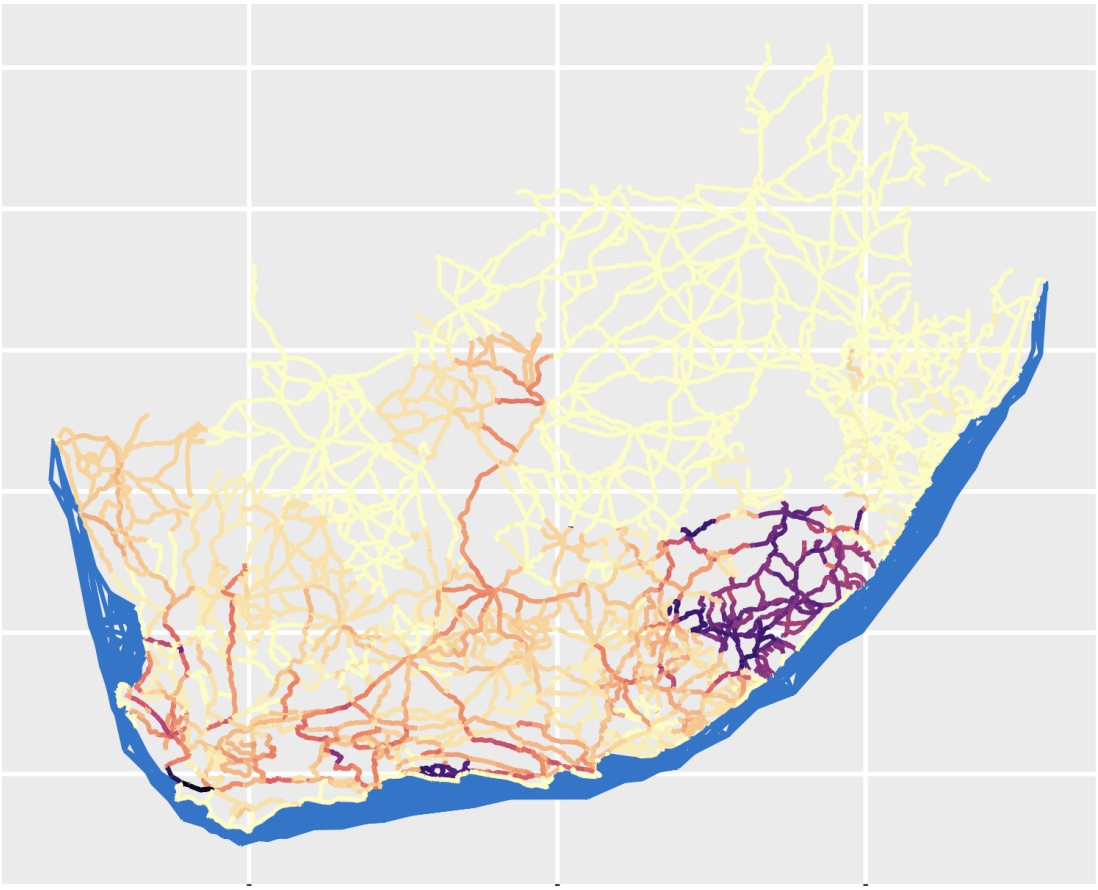
The first inequality shows that an increase in desegregation (such as market access to non-white populations) increases the marginal value to repress; in order to decrease the cost of repression the optimal value of n must be lower, which induces white settlers to identify

further from the British and closer to the Afrikaner $n^* < n_B$. The second inequality shows that the absence of desegregation (such as increases in market access to white populations) increases the marginal cost of not being British, increasing the optimal value of n which induces repression towards zero; the white settler will identify closer to the British and further away from the Afrikaner $n^* \approx n_B$.

White settlers' national identity choice is therefore driven by whether market access desegregates the local economy or not; for example, whether higher integration to non-white populations (or increases in market access to non-white agricultural locations) push desegregation at a faster pace than integration to white populations (or increases in economic opportunities to whites). The greater the desegregation, the higher the cognitive costs and lower the racial relative status, which increase the marginal value of institutionalized repression tilting identification towards the Afrikaner nation. The lower the desegregation, the higher the marginal costs of deviating from British norms and institutions, and more white settlers will identify with this the British nation.

B Additional Figures and Tables

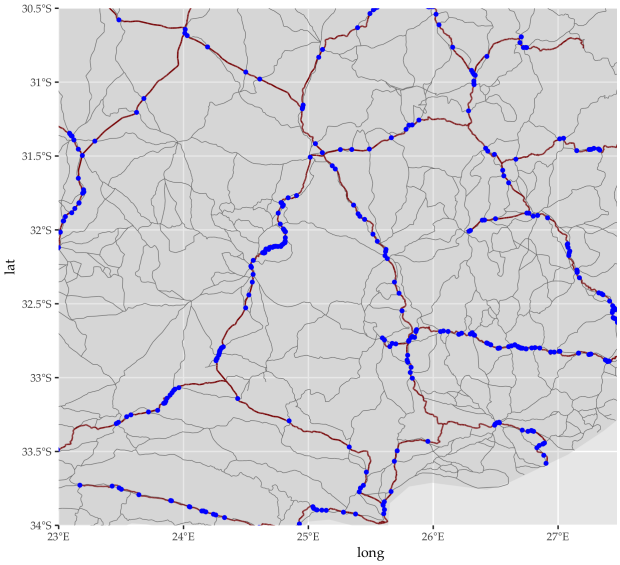
Figure B1: Road Construction Heat Map (1869-1909)



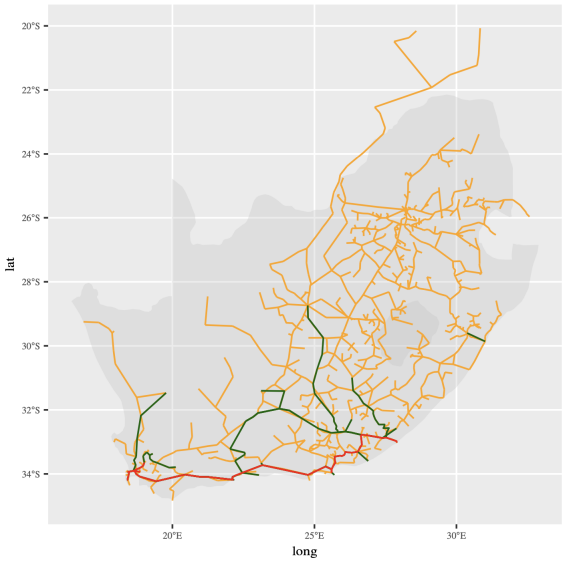
This figure depicts a heat map that aggregates the roads that were constructed on between 1869 and 1909. The darker the color the more years the road was worked on. The heterogeneity of this infrastructure across space is noticeable. Road construction was by far the highest in the Cape Colony, followed by Natal, and practically non-existing in the Orange Free State (OVS) and the Transvaal (ZAR). This is hardly surprising, as the two latter polities were hardly organized states with a state-capacity, while the two former were managed by the dominant British Empire. Invariant sea routes are included in blue.

Figure B2: Roads Intersections & Expansion of Telegraphs

(a) Intersections Zoom-in

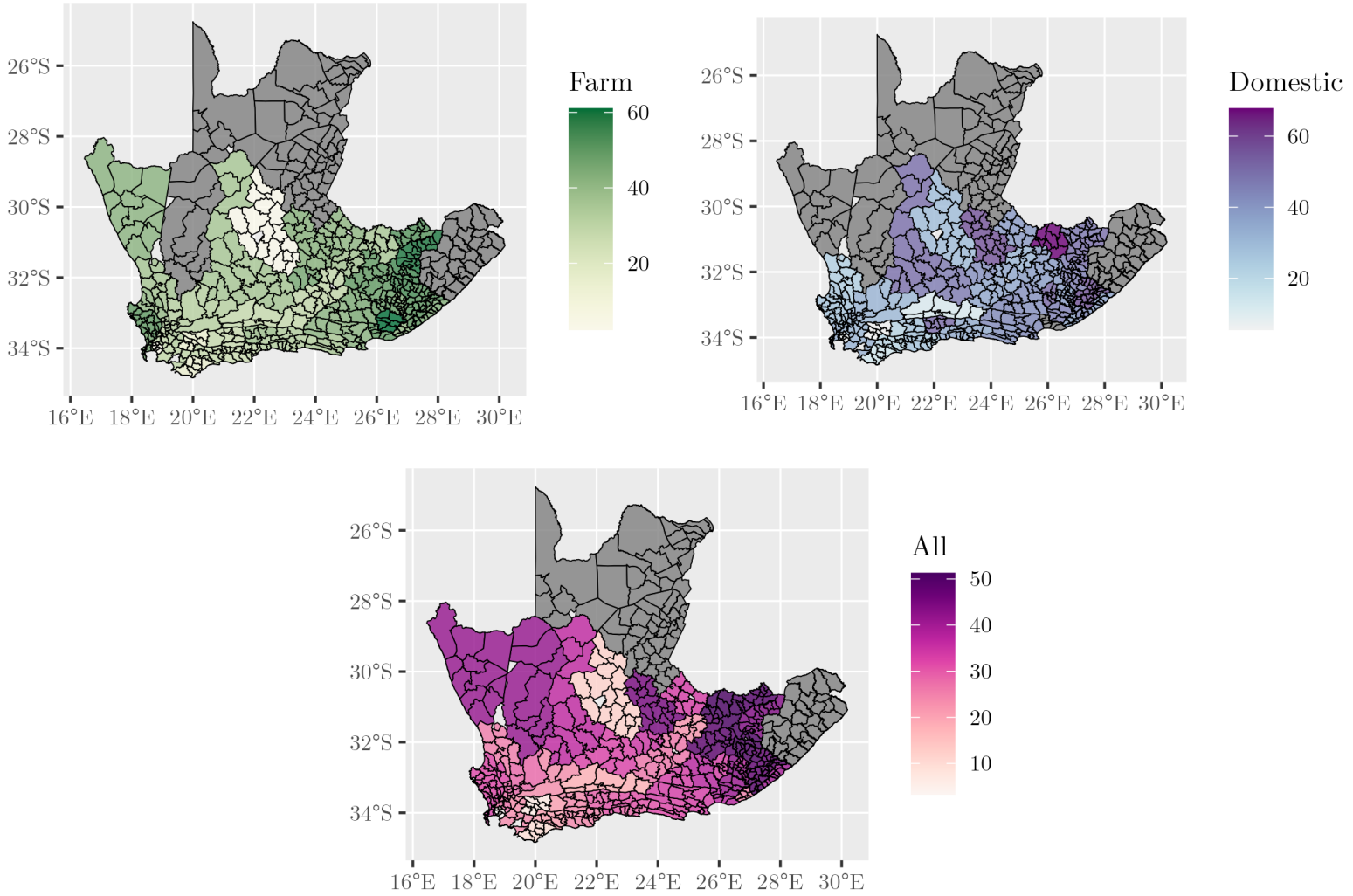


(b) Telegraph Expansion Example



The left panel shows an example telegraphs in 1872 (red), 1880 (green), and 1896 (orange). The right panel shows a zoomed-in map of rails in red, roads in grey, and intersections in blue.

Figure B3: Baseline Racial Wage Gaps (Cape of Good Hope)



These figures depict the racial wage gaps by set of occupations for the baseline years 1870-1874. The frontier is in gray. Wage gaps are reported at the district level; smaller polygons are sub-districts. The wage gap differences is considerable across locations.

Table B1: Estimated Impact of Market Access on Grassroots Organizations Meetings' Topics

	Politics	Economic	British	Africans
	(1)	(2)	(3)	(4)
MA (total)	0.080	0.292**	0.109*	0.093
	(0.092)	(0.139)	(0.062)	(0.140)
Observations	4,387	4,387	4,387	4,387
R^2	0.440	0.556	0.131	0.077
District fixed effects (107)	✓	✓	✓	✓
Year fixed effects (41)	✓	✓	✓	✓

The estimates report the impact of a one standard deviation greater increase in district market access on each of the dependent variables between 1870 and 1910 (yearly). All variables are in logarithms. All regressions include year and district fixed effects and standard errors are clustered at the district level. The first year has 70 districts; as the frontier expands districts reach 107.

Table B2: Estimated Impact of Market Access (total)
on MPs Occupations

	Farmer	Prof.	Comm.	Lawyer	Military
	(1)	(2)	(3)	(4)	(5)
MA (t)	0.331**	-0.186	-0.201	-0.093	-0.278**
	(0.150)	(0.187)	(0.136)	(0.153)	(0.118)
Observations	809	809	809	809	809
R^2	0.374	0.412	0.332	0.268	0.282
District fixed effects (106)	✓	✓	✓	✓	✓
Year fixed effects (9)	✓	✓	✓	✓	✓

The estimates report the impact of a one standard deviation greater increase in district market access on each of the dependent variables between 1870 and 1910 (every 5 years). All variables are in logarithms. All regressions include year and district fixed effects and standard errors are clustered at the district level. The first year has 70 districts; as the frontier expands districts reach 106.

Table B3: (*first*) Estimated Impact of Market Access on National Identity Choice
Heterogeneous Effects by Initial White Populations' Religions and Non-White Populations

(a) Dependent Variables: Children's Names

	Afrikaans (1)	English (2)	Dutch (3)
MA	0.067*	0.021	-0.049*
	(0.039)	(0.031)	(0.029)
MA × Protestants	-0.049*	-0.033	0.022
	(0.026)	(0.035)	(0.020)
MA × DRC	-0.0008	0.011	-0.021
	(0.027)	(0.036)	(0.019)
MA × Non-White	0.021	-0.009	0.007
	(0.018)	(0.018)	(0.008)
Observations	12,133	12,133	12,133
R^2	0.193	0.311	0.170
Sub-District FE (562)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: Newspaper Language

	Afrikaans (1)	English (2)	African (3)	Indian (4)
MA	0.174**	0.205***	0.176	0.019
	(0.070)	(0.060)	(0.114)	(0.036)
MA × Protestants	-0.033	0.038	0.239*	0.301
	(0.087)	(0.065)	(0.125)	(0.233)
MA × DRC	-0.038	-0.097	-0.309**	-0.434
	(0.099)	(0.064)	(0.143)	(0.318)
MA × Non-White	0.132**	0.061	0.061	0.156
	(0.061)	(0.056)	(0.089)	(0.111)
Observations	7,749	7,749	7,749	7,749
R^2	0.731	0.767	0.488	0.192
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. Panel B3a displays results for the share of children's first names of a given etymology. Market Access in interacted with number of baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations. Panel B3b displays results for the number of newspaper publications in different languages. Panel B3b are at the district level. Panels B3a are yearly panels. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B3: (*continued*) Estimated Impact of Market Access on National Identity Choice
Heterogeneous Effects by Initial White Populations' Religions and Non-White Populations

(a) Dependent Variables: Grassroots Orgs. Meetings

	Afrikaner B. (1)	Farmer Assoc. (2)	Other (3)
MA	-0.061	0.306**	0.184*
	(0.126)	(0.131)	(0.100)
MA × Protestants	-0.117	-0.016	0.042
	(0.079)	(0.040)	(0.071)
MA × DRC	0.144	-0.0007	-0.066
	(0.094)	(0.044)	(0.091)
MA × Non-White	0.070***	-0.005	0.006
	(0.025)	(0.022)	(0.026)
Observations	4,387	4,387	4,387
R^2	0.122	0.218	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

(b) Dependent Variables: MPs Last Names

	Dutch (1)	English (2)	non-English (3)
MA	0.161	-0.197	0.277***
	(0.172)	(0.140)	(0.096)
MA × Protestants	-0.303**	0.327***	-0.255**
	(0.122)	(0.103)	(0.110)
MA × DRC	0.238*	-0.286***	0.058
	(0.127)	(0.108)	(0.081)
MA × Non-White	0.039	-0.096**	-0.058
	(0.051)	(0.043)	(0.077)
Observations	809	809	809
R^2	0.373	0.484	0.538
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. Panel B4a displays results for the number of meetings for the pro-Afrikaner *Afrikaner Bond*, the pro-British *Farmers Protection Association*, and other agricultural organizations. Panel B4b displays results for the share of elected MPs' last names of a given etymology. Panel B4b is at the sub-district level. Panel B3b are at the district level. Panels B4b are yearly panels. Panel B4b is a five-yearly panel. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B4: Estimated Impact of Market Access on Baby Names (Interactions by Race)

	Afrikaans	English	Dutch
	(1)	(2)	(3)
MA (white)	-0.057	-0.015	0.092
	(0.153)	(0.126)	(0.120)
MA (non-white)	0.156	0.059	-0.141
	(0.148)	(0.125)	(0.117)
MA (white) × Non-White	-0.448**	-0.481**	0.154
	(0.189)	(0.207)	(0.193)
MA (non-white) × White	-0.385**	-0.401*	0.225
	(0.193)	(0.211)	(0.228)
MA (white) × White	0.360*	0.452**	-0.231
	(0.197)	(0.221)	(0.242)
MA (non-white) × Non-White	0.455**	0.413**	-0.148
	(0.184)	(0.199)	(0.185)
Observations	12,133	12,133	12,133
R^2	0.193	0.312	0.170
Sub-District FE (562)	✓	✓	✓
Year FE (41)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the share of children's first names of a given etymology.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B5: Estimated Impact of Market Access on Newspaper Language (Interactions by Race)

	Afrikaans	English	African	Indian
	(1)	(2)	(3)	(4)
MA (white)	-0.235	0.381**	0.257	0.088
	(0.172)	(0.173)	(0.253)	(0.124)
MA (non-white)	0.396**	-0.152	-0.076	-0.081
	(0.160)	(0.157)	(0.210)	(0.139)
MA (white) × Non-White	1.74	0.441	-0.647	-2.18
	(1.61)	(1.19)	(1.15)	(1.61)
MA (non-white) × White	1.36	0.117	-0.589	-2.15
	(1.50)	(1.08)	(1.13)	(1.61)
MA (white) × White	-1.84	-0.349	0.281	2.20
	(1.66)	(1.27)	(1.19)	(1.65)
MA (non-white) × Non-White	-1.20	-0.182	0.938	2.20
	(1.47)	(1.03)	(1.11)	(1.62)
Observations	7,749	7,749	7,749	7,749
R^2	0.732	0.768	0.486	0.172
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the number of newspaper publications in different languages.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B6: Estimated Impact of Market Access on Meetings (Interactions by Race)

	Afrikaner B.	Farmer Assoc.	Other
	(1)	(2)	(3)
MA (white)	-0.721***	0.840***	-0.024
	(0.258)	(0.301)	(0.238)
MA (non-white)	0.618***	-0.535**	0.184
	(0.222)	(0.209)	(0.261)
MA (white) × Non-White	2.65	-0.191	-0.157
	(2.48)	(2.18)	(1.60)
MA (non-white) × White	3.08	-0.163	-0.632
	(2.67)	(2.31)	(1.82)
MA (white) × White	-2.85	0.598	0.465
	(2.56)	(2.16)	(1.67)
MA (non-white) × Non-White	-2.85	-0.211	0.340
	(2.61)	(2.36)	(1.75)
Observations	4,387	4,387	4,387
R^2	0.125	0.220	0.388
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the number of meetings for the pro-Afrikaner *Afrikaner Bond*, the pro-British *Farmers Protection Association*, and other agricultural organizations.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B7: Estimated Impact of Market Access on MPs Names (Interactions by Race)

	Dutch	English	non-English
	(1)	(2)	(3)
MA (white)	-1.15***	0.580*	-0.695**
	(0.320)	(0.293)	(0.309)
MA (non-white)	1.39***	-0.837***	1.01***
	(0.362)	(0.314)	(0.332)
MA (white) × Non-White	3.82	2.25	-0.796
	(2.53)	(2.42)	(1.95)
MA (non-white) × White	4.10	2.81	-2.03
	(2.98)	(2.69)	(2.12)
MA (white) × White	-4.06	-2.11	1.19
	(2.61)	(2.47)	(1.98)
MA (non-white) × Non-White	-3.89	-3.02	1.45
	(2.94)	(2.66)	(2.11)
Observations	809	809	809
R^2	0.386	0.485	0.540
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the share of elected MPs' last names of a given etymology.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B8: Estimated Impact of Market Access on Baby Names (Interactions by Religion)

	Afrikaans	English	Dutch
	(1)	(2)	(3)
MA (Protestant)	0.136	0.301	-0.037
	(0.174)	(0.183)	(0.211)
MA (DRC)	-0.076	-0.148	-0.021
	(0.176)	(0.183)	(0.208)
MA (non-white)	0.036	-0.099	-0.013
	(0.082)	(0.078)	(0.079)
MA (DRC) × DRC	-0.330*	0.216	-0.226
	(0.176)	(0.181)	(0.211)
MA (Protestant) × DRC	0.225	-0.099	0.161
	(0.146)	(0.157)	(0.189)
MA (non-white) × DRC	0.069	-0.096	0.043
	(0.072)	(0.066)	(0.063)
Observations	14,083	14,083	14,083
R^2	0.190	0.302	0.160
Sub-District FE (687)	✓	✓	✓
Year FE (41)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the share of children's first names of a given etymology.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B9: Estimated Impact of Market Access on Newspaper Language (Interactions by Religion)

	Afrikaans	English	African	Indian
	(1)	(2)	(3)	(4)
MA (Protestant)	-0.070	0.143	-0.216	-0.604
	(0.288)	(0.247)	(0.326)	(0.427)
MA (DRC)	-0.085	0.101	0.288	0.442
	(0.280)	(0.257)	(0.301)	(0.355)
MA (non-white)	0.386**	0.023	0.157	0.184
	(0.163)	(0.104)	(0.179)	(0.135)
MA (DRC) × DRC	0.325	0.489**	0.253	-0.334
	(0.237)	(0.229)	(0.319)	(0.285)
MA (Protestant) × DRC	-0.431*	-0.607**	-0.586*	0.428
	(0.233)	(0.242)	(0.344)	(0.341)
MA (non-white) × DRC	0.084	0.038	0.200	-0.074
	(0.118)	(0.100)	(0.179)	(0.058)
Observations	7,749	7,749	7,749	7,749
R^2	0.728	0.769	0.488	0.175
District FE (189)	✓	✓	✓	✓
Year FE (41)	✓	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the number of newspaper publications in different languages.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B10: Estimated Impact of Market Access on Meetings (Interactions by Religion)

	Afrikaner B.	Farmer Assoc.	Other
	(1)	(2)	(3)
MA (Protestant)	-0.426	-1.09*	-0.105
	(0.275)	(0.624)	(0.396)
MA (DRC)	-0.459*	1.01**	-0.022
	(0.253)	(0.395)	(0.187)
MA (non-white)	0.691**	0.403	0.262
	(0.319)	(0.405)	(0.412)
MA (DRC) × DRC	0.101	0.124	0.207
	(0.202)	(0.122)	(0.187)
MA (Protestant) × DRC	-0.235	0.196	0.147
	(0.146)	(0.131)	(0.147)
MA (non-white) × DRC	0.147	-0.207**	-0.282
	(0.174)	(0.097)	(0.175)
Observations	4,387	4,387	4,387
R^2	0.124	0.222	0.389
District FE (107)	✓	✓	✓
Year FE (41)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the number of meetings for the pro-Afrikaner *Afrikaner Bond*, the pro-British *Farmers Protection Association*, and other agricultural organizations.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

Table B11: Estimated Impact of Market Access on MPs Names (Interactions by Religion)

	Dutch	English	non-English
	(1)	(2)	(3)
MA (Protestant)	-1.08**	-0.684	-0.424
	(0.462)	(0.426)	(0.400)
MA (DRC)	-0.334	0.786***	-0.209
	(0.244)	(0.238)	(0.293)
MA (non-white)	1.37***	-0.285	0.713**
	(0.356)	(0.326)	(0.304)
MA (DRC) × DRC	-0.859***	0.233	0.012
	(0.290)	(0.188)	(0.267)
MA (Protestant) × DRC	1.18**	-0.302	1.23***
	(0.451)	(0.260)	(0.372)
MA (non-white) × DRC	0.036	-0.038	-0.947***
	(0.275)	(0.163)	(0.193)
Observations	809	809	809
R^2	0.395	0.490	0.550
District FE (106)	✓	✓	✓
Year FE (9)	✓	✓	✓

This table reports estimates of the impact of a one standard deviation greater increase in total market access on each set of dependent variables using equation 1. displays results for the share of elected MPs' last names of a given etymology.. Market Access to baseline white Protestant churches members, white Dutch Reformed Church members, and non-white populations are interacted with total number of baseline Dutch Reformed Church members. All variables are in logarithms. All regressions include year and sub-district/district fixed effects and standard errors are clustered at the sub-district/district level, the level of the treatment.

C Selected Primary Sources on the Cape of Good Hope Colony

C.1 National Archives of South Africa: Cape Town Archive Repository (KAB)

C.1.1 Census District Maps (1904). Compiled from maps filed in the Surveyor General's Office

:

- Map of the division of Aberdeen showing field-cornetcies, farms with grant numbers, etc. on which are indicated 1904 census particulars. [map]. M3/2038. 1902.
- Divisional map of Albany showing {...}. [map]. M3/2039. 1890.
- Divisional map of Albert showing {...}. [map]. M3/2040. 1902.
- Divisional map of Alexandria showing {...}. [map]. M3/2041. 1899.
- Divisional map of Aliwal North showing {...}. [map]. M3/2043. 1904.
- Divisional map of Barkly East. [map]. M2/1922. 1900.
- Divisional map of Barkly East {...}. [map]. M3/2045. 1899.
- Divisional map of Barkly West showing {...}. [map]. M3/2046. 1899.
- Divisional map of Bathurst s{...}. [map]. M3/2047. 1890.
- Divisional map of Beaufort West {...}. [map]. M3/2048. 1899.
- Divisional map of Bedford {...}.. [map]. M3/2049. 1900.
- Divisional map of Bredasdorp {...}. [map]. M3/2050. 1901.
- Map of the division of Britstown{...}. [map]. M3/2051. 1901.
- Map of the district of Butterworth{...}. [map]. M/2155. 1903.
- Divisional map of Caledon {...}. [map]. M3/2052. 1904.
- Divisional map of Calvinia {...} [map]. M3/2053. 1900.
- Divisional map of Carnarvon {...}. [map]. M3/2054. 1900.
- Divisional map of Cathcart {...}. [map]. M3/2055. 1901.
- Divisional map of Ceres {...}.. [map]. M3/2056. 1901.
- Divisional map of Clanwilliam {...}. [map]. M3/2057. 1900.
- Divisional map of Colesberg{...}. [map]. M3/2058. 1900.
- Divisional map of Gordonia {...}. [map]. M3/2059. 1900.
- Divisional map of Cradock {...}. [map]. M3/2060. 1901.
- General plan of East London {...}. [map]. M3/2062. 1904.
- Map of the division of East London {...}. [map]. M3/2061. 1898-1889.
- Map of the district of Elliotdale {...}. [map]. M3/2149. 1904.
- Map of the district of Elliot Slang River {...}. [map]. M3/2148. 1902.
- Divisional map of Fort Beaufort {...}. [map]. M3/2064. 1900.
- Divisional map of Fraserburg {...}. [map]. M3/2065. 1901.
- Divisional map of George {...}. [map]. M3/2066. 1900.
- Map of the division of Graaff-Reinet {...}. [map]. M3/2068. 1900.
- Map of the division of Hanover {...}. [map]. M3/2070. 1899.
- Divisional map of Hay {...}. [map]. M3/2071. 1900.
- Divisional map of Herbert {...}. [map]. M3/2072. 1900.
- Map of the division of Hopetown {...}. [map]. M3/2074. 1902.

- Divisional map of Humansdorp {...}. [map]. M3/2075. 1901.
- Map of the district of Idutywa {...}. [map]. M3/2133. 1904.
- Map of the district of Kentani {...}. [map]. M3/2156. 1902.
- Map of the district of Kentani. {...}. [map]. M4/1355. 1902.
- Map of the district of Kentani. {...}. [map]. M4/1356. 1902.
- Divisional map of Kimberley {...}.. [map]. M3/2080. 1899.
- Divisional map of King William's Town, {...}. [map]. M1/250. 1900.
- Divisional map of King William's Town, {...}. [map]. M1/251. 1900.
- Divisional map of King William's Town, {...}.. [map]. M1/252. 1900.
- Divisional map of King William's Town, {...}. [map]. M1/253. 1900.
- Divisional map of King William's Town, {...}.. [map]. M1/254. 1900.
- Divisional map of King William's Town, {...}. [map]. M1/255. 1900.
- Divisional map of King William's Town, {...}.. [map]. M1/256. 1900.
- Divisional map of King William's Town, {...}.[map]. M1/257. 1900.
- Divisional map of Knysna {...}. [map]. M3/2083. 1901.
- Divisional map of Komgha {...}. [map]. M3/2084. 1902.
- Divisional map of Ladysmith {...}. [map]. M3/2085. 1900.
- Map of the division of Maclear {...} [map]. M3/2513. 1902.
- Map of the district of Maclear [map]. M3/2138. 1902.
- Map of the district of Maclear. {...} [map]. M4/1352. 1902.
- Map of the district of Maclear. {...} [map]. M4/1353. 1902.
- Map of the district of Maclear. {...}. [map]. M4/1354. 1902.
- Divisional map of Mafeking, {...} [map]. M3/2168. 1900.
- Divisional map of Malmesbury {...} [map]. M3/2086. 1900.
- Map of the district of Matatiele {...}. [map]. M3/2087. 1900.
- Divisional map of Molteno {...}. [map]. M3/2089. 1900.
- Divisional map of Mossel Bay {...} [map]. M3/2090. 1901.
- Map of the district of Mount Ayliff {...} [map]. M3/2140. 1903.
- Map of the district of Mount Currie{...} [map]. M3/2141. 1902.
- Map of the district of Mount Fletcher {...} M3/2142. 1904.
- Map of the district of Mount Frere, East Griqualand {...} [map]. M3/2143. 1903.
- Map of the Mqanduli district {...}. [map]. M3/215. 1903.
- Map of the division of Murraysburg {...} [map]. M3/2091. 1902.
- Divisional map of Namaqualand {...} [map]. M3/2092. 1904.
- Divisional map of Namaqualand {...} [map]. M3/2093. 1904.
- Divisional map of Oudtshoorn {...} [map]. M3/2094. 1900.
- Divisional map of Paarl {...} [map]. M3/2096. 1904.
- Divisional map of Peddie {...} [map]. M3/2097. 1902.
- Divisional map of Philipstown {...} [map]. M3/2098. 1904
- Divisional map of Piketberg {...} [map]. M3/2099. 1900.
- Plan of Port Elizabeth {...}. [map]. M3/2100. 1902.
- Plan of part of Port Elizabeth {...}. [map]. M3/2101. 1903.

- Divisional map of Port Elizabeth and Uitenhage {...} [map]. M3/2103. 1904.
- Divisional map of Prieska {...}.. [map]. M3/2104. 1900.
- Plan of Queenstown. Census 1904. [map]. M3/2110. 1903.
- Plan of Queenstown. Census 1904 [map]. M3/2101. 1904.
- Plan of Queenstown. Census 1904 [map]. M3/2108. 1904.
- Plan of part of Queenstown {...} [map]. M3/2109. 1904.
- Map of the district of Qumbu {...} [map]. M3/2144. 1904.
- Street plan of Richmond {...} [map]. M3/2112. 1903.
- Divisional map of Richmond. {...} [map]. M2/281. 1901.
- Divisional map of Richmond {...} [map]. M3/3289. 1901.
- Divisional map of Richmond {...} [map]. M3/3290. 1901.
- Map of the division of Richmond. {...} [map]. M3/4804. 1901.
- General plan of the village of Riversdale{...} [map]. M3/2114. 1900.
- Map of the division of Robertson {...} [map]. M3/2115. 1901.
- Rough plan showing the Municipal limits of the village of Robertson, {...}. [map]. M3/2116. 1904.
- Divisional map of Somerset East {...} [map]. M3/2117. 1900.
- Map of the Stellenbosch division{...} [map]. M3/2118. 1899.
- Map of the division of Steynsburg {...} [map]. M3/2119. 1901.
- Divisional map of Stockenstrom {...} [map]. M3/2120. 1902.
- Divisional map of Stutterheim {...} [map]. M3/2121. 1901.
- Divisional map of Sutherland{...}. [map]. M3/2122. 1900.
- Divisional map of Swellendam {...}. [map]. M3/2123. 1901.
- Divisional map of Tarka {...}. [map]. M3/2125. 1901.
- Divisional map of Tulbagh{...}[map]. M3/2126. 1900.
- Divisional map of Uitenhage {...}. [map]. M3/2127. 1900.
- Map of the Umtata district {...}. [map]. M3/2153. 1904.
- Map of the division of Uniondale{...}[map]. M3/2128. 1901.
- Map of the division of Van Rhynsdorp {...}. [map]. 1904 M3/2129. 1900.
- Divisional map of Victoria East {...} [map]. 1900 M3/2130 1900.
- Divisional map of Victoria West{...}. [map]. M3/2131.1900.
- Part of divisional map of Vryburg {...}.. [map]. M3/2169. 1900.
- Part of divisional map of Vryburg {...}. [map]. M3/2170. 1900
- Part of divisional map of Vryburg {...}. [map]. M3/2171. 1900.
- Part of divisional map of Vryburg {...}. [map]. M3/2172. 1902.
- Divisional map of Willowmore{...}. [map]. M3/2132. 1900.
- Rough plan of the Municipality of Willowmore {...}.. [map]. M3/2134. 1904.
- Map of the district of Willowvale{...}.[map]. M3/2159. 1902.
- Divisional map of Wodehouse {...}.. [map]. M3/2135. 1904.
- Divisional map of Worcester{...}.. [map]. M3/2136. 1901.
- Map of the district of Tsolo {...}. M3/2145.1904.
- Map of the district of Umzimkulu {...}. [map]. M3/2146. 1902.
- Map of the district of Elliot (Xalanga) {...}. [map]. M3/2147. 1890.
- Map of the district of Xalanga {...}. [map]. M3/2154. 1902.

- Map of Pondoland {...}. [map]. M3/2161. 1902.
- Map of Pondoland {...}.. [map]. M3/2162. 1902.
- Map of Pondoland {...}. [map]. M3/2163. 1902.
- Map of Pondoland showing Lusikisiki. {...}. [map]. M3/2164. 1902.
- Map of Pondoland in Ngqeleni. {...} [map]. M3/2165. 1902.
- Map of Pondoland.showing Port St. Johns. {...} [map]. M3/2166. 1902.
- Map of Pondoland Tabankulu. {...} [map]. M3/2167. 1902.

C.1.2 Blue Books and Statistical Registers (1869-1909). British Foreign & Commonwealth Office:

- Blue Book for the Colony of the Cape of Good Hope, 1869 (-1885). Cape Town: Saul Solomon & Co., Steam Printing Office, 49 & 50 St George's-Street [publishers vary by year]. 1870-1886.
- Statistical Register of the Colony of the Cape of Good Hope for the Year 1886 (-1909). Cape Town: W. A. Richards and Sons, Government Printers, Castle Street [publishers vary by year]. 1887-1910.

C.2 Stellenbosch University

C.2.1 Railway Records

- Report of The General Manager of Railways for the Year 1909. Cape Town: Cape Government Railways. 1909.

References

Acemoglu, D., & Robinson, J. (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Akerlof, G., & Kranton, R. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715-53.

Akerlof, G., & Kranton, R. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, *40*(4), 1167-1201.

Akerlof, G., & Kranton, R. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.

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

Algan, Y., Mayer, T., & Thoenig, M. (2022). The economic incentives of cultural transmission: Spatial evidence from naming patterns across France. *The Economic Journal*.

Alvarez, E., Bogart, D., Satchell, M., Shaw-Taylor, L., & You, X. (2013). Railways and structural change: evidence from industrializing britain. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *46*(3), 175-191.

Anderson, B. (1983). *Imagined communities: reflections on the origin and spread of nationalism*. Verso.

Andersson, D., Berger, T., & Prawitz, E. (2020). Making a market: Infrastructure, integration and the rise of innovation. *Working Paper.*.

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(337-341).

Aschenborn, H. (1983). *A list of south african newspapers 1800-1982, whith library holdings* (S. L. Pretoria, Ed.).

Assouad, L. (2021). Charismatic leaders and nation-building. *Working Paper*.

Atkin, D., Colson-Sihra, E., & Shayo, M. (2021). How do we choose our identity? A revealed preference approach using food consumption. *Journal of Political Economy*, *129*(4), 1193 -1251.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion,. *Proceedings of the 21st International World Wide Web Conference,*(519-528).

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, *145*(102442).

Barth, F. (1969). Ethnic groups and boundaries: The social organization of culture difference. In F. Barth (Ed.), (p. 9-38). Little Brown.

Beckert, S. (2015). *Empire of cotton*. Penguin Books.

Berger, T., & Enflo, K. (2014). Locomotives of local growth: The short- and long-term impact of railroads in sweden. *Lund Papers in Economic History*.

Bisin, A., & Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*.

Bisin, A., & Verdier, T. (2017). On the joint evolution of culture and institutions. *CEPR Discussion Papers*(12000).

Blanc, G., & Kubo, M. (2021). Schools, language, and nations: Evidence from a natural experiment in france. *Working Paper*.

Blouin, A., & Mukand, S. (2019). Erasing ethnicity? propaganda, nation building, and identity in rwanda. *Journal of Political Economy*, *127*(3), 1008-1062.

Bogart, D., You, X., Alvarez-Palau, E., Satchell, M., & Shaw-Taylor, L. (2018). Railways, growth and spatial reorganisation: Evidence from nineteenth century england and wales. *Working Paper.*.

Bonacich, E. (1972). A theory of ethnic antagomism: the split labor market. *American Sociological Review*, *37*, 547-59.

Bonacich, E. (1973). A theory of middleman minorities. *American Sociological Review*, *38*, 583-94.

Bonacich, E. (1976). Advanced capitalism and black/white relations. *American Sociological Review*, *41*, 31-51.

Boruskyak, K., & Hull, P. (2022). Non-random exposure to exogenous shocks. *Working Paper*.

Bundy, C. (1979). *The rise and fall of the african peasantry*. University of California Press.

Cantoni, D., Chen, Y., Yang, D., Yuchtman, N., & Zhang, Y. (2017). Curriculum and ideology. *Journal of Political Economy*, *125*(2).

Caprettini, B., & Voth, H.-J. (2020). Rage against the machines: Labor-saving technology and unrest in industrializing england. *American Economic Review: Insights (forthcoming)*.

Chesterton, G. (1917). *A short history of England*. Digibooks OOD.

Davenport, T. (1960). *The Afrikaner Bond 1880-1900* (Unpublished doctoral dissertation). University of Cape Town.

Dehdari, S., & Gehring, K. (2022). The origins of common identity: Evidence from Alsace-Lorraine. *American Economic Journal: Applied Economics*.

Dell, M., & Querubin, P. (2017). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, *133*(2), 701-764.

Depetris-Chauvin, E., Durante, R., & Campante, F. (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review*, *110*(5), 1572-1602.

Deumert, A. (2004). *Language standardization and language change: The dynamics of cape dutch*. John Benjamins Publishing.

Deutsch, K. (1953). *Nationalism and social communication, second edition: An inquiry into the foundations of nationality*. The M.I.T. Press.

Dominic Rohner, F. Z., Mathias Thoenig. (2013). Seeds of distrust: Conflict in Uganda. *Journal of Economic Growth*.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899-934.

Donaldson, D., & Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, *131*(2), 799-858.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting behavior and the internet. *American Economic Review*, *104*(7), 2238-2265.

Farley, J. (1988). *Majority-minority relations*. Prentice Hall.

Fernández, R. (2010). Does culture matter? *NBER Working Papers*(16277).

Fishlow, A. (1965). American railroads and the transformation of the ante-bellum economy. *Harvard University Press*.

Fogel, R. (1964). *Railroads and american economic growth*. Johns Hopkins Press Baltimore.

Fouka, V. (2019). How do immigrants respond to discrimination? the case of germans in the us during world war i. *American Political Science Review*, *113*(2), 405-422.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in u.s. schools after world war i. *Review of Economic Studies*, *87*, 204-239.

Fourie, J., & Herranz-Loncan, A. (2015). Growth (and segregation) by rail: How the railways shaped colonial south africa. *ERSA working paper 539*.

García-Jimeno, C., Iglesias, A., & Yildirim, P. (2018). Women, rails, and telegraphs: An

empirical study of information diffusion and collective action‘. *NBER Working paper*.

Gellner, E. (1983). *Nations and nationalism*. Cornell University Press.

Gilliomee, H. (2009). *The afrikaners* (U. of Virginia Press, Ed.).

Gorodnichenko, Y., & Roland, G. (2011). Which dimensions of culture matter for long-run growth? *The American Economic Review,*, *101*(2), 492-498.

Grundlingh, M. (1945). *The parliament of the Cape of Good Hope. with special refernce to party politics, 1872 to 1910* (Unpublished doctoral dissertation). University of Stellenbosch.

Hannan, M. (1979). National development in the world system: Educational, economic, and political change, 1950-1970. In J. Meyer & M. Hannan (Eds.), (p. 253-275). University of Chicago Press.

Harari, Y. (2015). *Sapiens : a brief history of humankind*. Harper.

Hechter, M. (2000). *Containing nationalism*. Oxford University Press.

Heeringa, W., & DeWet, F. (2015). The origin of the afrikaans pronunciation: a comparison to west germanic languages and dutch dialects. *Stellenbosch Papers in Linguistics Plus*, *47*.

Heese, J., Nienaber, G., & Nienaber, C. (1975). *Families, familiename en familiewapens*. Tafelberg.

Henrich, J. (2020). *The weirdest people in the world: How the west became psychologically peculiar and particularly prosperous* (S. Farrar & Giroux, Eds.). Farrar,.

Hobsbawm, E. (1990). *Nations and nationalism since 1780: Programme, myth, reality*. Cambridge University Press.

Hoetink, H. (1975). Ethnicity and resource competition in plural societies. In L. Despres. (Ed.), (p. 9-25). De Gruyter Mouton.

Hornbeck, R., & Rotemberg, M. (2019). Railroads, reallocation, and the rise of american manufacturing. *NBER Working Paper (26594)*.

Hornung, E. (2015). Railroads and growth in prussia. *European Economic Association*, *13*(4), 699-736.

Hroch, M. (1985). *Social preconditions of national revival in europe: A comparative analysis of the social composition of patriotic groups among the smaller european nations* (F. Ben, Ed.). Cambridge University Press.

Jansen, C., Schreuder, R., & Neijt, A. (2007). The influence of spelling conventions on

perceived plurality in compounds. *Written Language and Literacy*, *10*(2), 185-194.

Jedwab, R., Kerby, E., & Moradi, A. (2015). History, path dependence and development: Evidence from colonial railroads, settlers and cities in kenya. *The Economic Journal*.

Jenne, E. (2004). A bargaining theory of minority demands: Explaining the dog that did not bite in 1990s yugoslavia. *International Studies Quarterly*, *48*(4), 729-754.

Katz, O. (2018). Railroads, economic development, and the demographic transition in the united states. *Working Paper.*.

Krige, J. (1936). *Die franse familiename in suid-afrika [van vóór 1800] etimologies verklaar*. Pretoria, Uitgegee vir die Universiteit van die Witwatersrand.

Laitin, D. (2007). *Nations, states, and violence*. Oxford University Press.

Maré, G. (1992). *Ethnicity and politics in south africa*. Zed Books.

Markoff, J. (1996). *Waves of democracy: Social movements and political change* (U. of Pittsburgh, Ed.). Pine Forge Press.

Maystre, N., Olivier, J., Thoenig, M., & Verdier, T. (2014). Product-based cultural change: Is the village global? *Journal of International Economics*, *92*, 212-230.

McCracken, J. (1967). *The cape parliament, 1854-1910*. Clarendon P.

Melander, E. (2020). Individual mobility and social mobilisation: Evidence from swedish grassroots movements. *Working paper*.

Molinder, J., Karlsson, T., & Enflo, K. (2019). More power to the people: Electricity adoption, technological change and social conflict. *Lund Papers in Economic History*.

Moodie, T. (1975). *The rise of afrikanerdom: Power, apartheid, and the afrikaner civil religion*. University of California Press.

Morton, F. (1992). Slave-raiding and slavery in the western Transvaal after the sand river convention. *African Economic History*, *1*(20), 99-118.

Munashinghe, V. (2018). The international encyclopedia of anthropology. In H. Callan (Ed.), (chap. Ethnicity in Anthropology). John Wiley & Sons.

Nagel, J. (1995). Resource competition theories. *American Behavioral Scientist*, *38*(3), 442-458.

Nagel, J., & Olzak, S. (1982). Ethnic mobilization in new and old states: An extension of the competition model. *Social Problems*, *30*(2), 127-143.

Nielsen, F. (1985). Toward a theory of ethnic solidarity in modern societies. *American Sociological Association*.

Olzak, S. (1992). *The dynamics of ethnic competition and conflict*. Stanford University Press.

Perlman, E. (2016). Connecting the periphery: Three papers on the developments caused by spreading transportation and information networks in the nineteenth century us. *Dissertation*.

Pomeranz, K. (2000). *The great divergence: China, europe, and the making of the modern world economy*. Princeton University Press.

Raidt, E. (1985). Historiese perspektief op die normering van afrikaans. *African Journal*, *1985*(10).

Samuels, D., & Thomson, H. (2020). Lord, peasant... and tractor? agricultural mechanization, moore's thesis, and the emergence of democracy. *Cambridge University Press*.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review*, *103*(2), 147-174.

Shayo, M. (2020). Social identity and economic policy. *Annual Review of Economics*, *12*, 355-89.

Startz, M. (2017). The value of face-to-face: Search and contracting problems in nigerian trade. *Working Paper*.

Tarrow, S. (1994). *Power in movement: Social movements and contentious politics*. Cambridge University Press.

Tilly, C. (1986). *The contentious french*. Belknap Press.

Toynbee, A. (1959). Optima. *History's Warning to Africa*.

Weber, E. (1976). *Peasants into frenchmen : the modernization of rural France, 1870-1914*. Stanford University Press.

Wimmer, A. (2008). The making and unmaking of ethnic boundaries: A multilevel process theory. *American Journal of Sociology*, *113*(4), 970-1022.