I

Introduction

1.1 Endogenous and Exogenous Ethnicity

A number of years ago I had the opportunity to sit down for dinner at a nice restaurant in Kampala, the capital of Uganda, with one of the country’s most prominent journalists, Andrew Mwenda, and one of its most prominent academics. Much of the conversation was dominated by Andrew, whose extraordinary knowledge of Ugandan political economy comes across in his magazine The Independent as well as his academic writings (cf. Mwenda, 2007; Tangri & Mwenda, 2008; among many others). Amid a discussion of my research interests into ethnic identity and the politics of land in Uganda, Andrew at one point described how a former Army captain had become a minister in President Yoweri Museveni’s government, got rich and then bought a large plot of land back in his home region of western Uganda. Puzzled, I asked why he invested his money in this way, as opposed to opening a Swiss bank account, buying foreign property in Dubai or investing abroad in some other way. Andrew’s vehement response was simple: ‘because this is still an agrarian country, my friend!’ According to him, the importance of owning rural land for Ugandan elites trumped their interest in financial security via foreign holdings due to the continued rural nature of the Ugandan economy, and thus explained the preponderance of Ugandan elite investment in local land holdings.

This conversation inspired me to consider the radically different nature of political and economic incentives in rural, agrarian dominated societies like Uganda as compared to urban, industrialized countries that are preponderantly in the Global North. More specifically, I began to consider how the highly fragmented nature of ethnic politics in Uganda, which I had already explored as regards land ownership and internal migration (Green, 2006, 2007), were a consequence of an economy based on agricultural production. The result of
Introduction

roughly a decade of subsequent research is this book, where I argue that industrialization creates incentives for individuals to re-identify ethnically and assimilate from small, more narrowly defined rural tribal identities to larger, more urban-focussed ethnic groups, and that this process is a consequence of the declining economic importance of control over rural land. To understand this process – and the intricacies of my argument – I momentarily take a step back to explain the broader origins of my story by turning again to Africa.

Despite high expectations as countries across Africa became independent of colonial rule in the 1950s and 1960s, the continent soon began to experience widespread problems with state collapse, conflict and underdevelopment in the last three decades of the twentieth century. As such social scientists searched for explanations as to why African states underperformed: particularly influential in this regard was a seminal article by William Easterly and Ross Levine that argued that ‘Africa’s Growth Tragedy’ was a consequence of its high level of ethnic fractionalization (Easterly & Levine, 1997). This paper – which today remains by far the most-cited among all of Easterly’s work – has led to a cottage-industry of examining the effects of ethnic fractionalization on various outcomes across the world, including economic growth and unemployment (Churchill & Smyth, 2017; Feldmann, 2012; Gören, 2014; Hjort, 2014; Montalvo & Reynal-Querol, 2005a; Spolaore & Wacziarg, 2009), civil wars (Bleaney & Dimico, 2011; Cederman & Girardin, 2007; Collier & Hoeffler, 2004; Costalli, Moretti, & Pischedda, 2017; Fearon & Laitin, 2003; Haynes, 2016; Hegre & Sambanis, 2006; Manotas-Hidalgo, Pérez-Sebastián & Campo-Bescós, 2021; Miguel, Satyanath, & Sergenti, 2004; Montalvo & Reynal-Querol, 2005b; Walter, 2006; Wegenast & Basedau, 2014), public policy and public goods provision (Alesina, Gennaioli & Lovo, 2019; Beach & Jones, 2017; Churchill, Ocloo & Siavor-Robertson, 2017; Habyarimana, Humphreys & Posner, et al., 2007; Lieberman, 2007; Miguel & Gugerty, 2005), and political party formation (W. R. Clark & Golder, 2006; Harbers, 2010), among other phenomena.

While much of this literature has been useful in understanding underdevelopment in the developing world and in Africa in particular, this body of scholarship largely treats ethnicity as fixed. In one sense, this assumption is non-problematic according to the most basic understanding of ethnicity in the literature, which is that ethnic groups are communities based on the idea that members share some form of common descent (Chandra, 2006; Fearon, 2003; Horowitz, 2000; A. D. Smith, 1991; M. Weber, 1978 [1922]). The question about whether or not members of a given group actually do share common ancestors is largely irrelevant; what matters instead is that a belief in common descent is prevalent enough that it binds people together in a single community. If this belief is common and unchanging, then an assumption that ethnicity is fixed over time is unproblematic.

However, the assumption of fixed ethnic identities has become increasingly problematic due to a shift since the 1960s towards more quantitative
1.2 The Argument

methodologies in the field of political economy, which has meant both a greater need for datasets which assign ethnic identities to people as well as the need to assume for game-theoretical purposes the existence of stable ethnic identities which inform players’ choices. This conception of ethnicity, where each person has one and only one ethnic identity which is inherited and thus cannot be changed, is known as primordialism. These primordialist assumptions have long been dominant in the political science and economics literature on ethnicity, such that scholars can readily assume that, to take one example, ethnic diversity is an ‘exogenously determined social state’ (Ordeshook & Shvetsova, 1994, p. 108). To take some examples, a notable amount of recent scholarship uses country-year panel data to examine the effects of ethnicity on a number of outcomes, with different annual measurements for conflict, democracy, GDP, population and price indices over multiple decades (and often up to a half-century), but with fixed measures of ethnic diversity over the same time span (Bleaney & Dimico, 2011; Gören, 2014; Haynes, 2016; Manotas-Hidalgo et al., 2021). In almost all of these studies ethnicity is implicitly assumed to be exogenous to the other variables of interest, with the sole exception of migration, while in the others the authors acknowledge the potential endogeneity of ethnicity but leave its causes and consequences for further investigation.

1.2 THE ARGUMENT

The goal of this book is thus to show that ethnicity is not in any way exogenous to changes in society, and, more specifically, that industrialization causes assimilation. My goal in this section is thus to lay out my argument in detail, which I can do in a series of steps. The first such step is as follows:

1. People hold multiple concentric ethnic identities, such that they can choose to emphasize one or another of these identities in a given context.

As noted above, historically scholars have assumed that ethnicity is primordial, such that individuals hold one and only one fixed and unchangeable ethnic identity. This attitude is arguably based on a ‘folk’ theory of primordial ethnicity held by non-academics (Gil-White, 1999), and goes back

---

1 Cf. Mauro (1995, p. 692), who similarly writes that ‘I assume that the extent to which countries are fractionalized along ethnolinguistic lines is exogenous and unrelated to economic variables.’


3 For instance, Miguel and Gugerty (2005, p. 2337) assume that levels of ethnic diversity in western Kenya ‘are largely the product of [pre-colonial] historical accident rather than recent migration’, thereby foreclosing other explanations for ethnic diversity.

4 Cf. Cederman and Girardin (2007, p. 176), who note that their use of ethnicity as an exogenous variable in their analysis of civil wars ‘does not mean that we believe that identities are primordially given’.
Introduction

decades or even centuries in the study of ethnicity and identity more broadly, as seen in the writings of the Jewish-American philosopher Horace Kallen, who wrote over a century ago that ‘Men may change their clothes, their politics, their wives, their religions, their philosophies, [but] they cannot change their grandfathers. Jews or Poles or Anglo-Saxons, in order to cease being Jews or Poles or Anglo-Saxons, would have to cease to be’ (Kallen, 1996 [1915], p. 91).

Recently this assumption has been built into the exercise of quantifying ethnicity in cross-national databases, whereby scholars have assigned people to an ethnic identity without any qualifications about the potential for this identity to vary across time (among others, cf. Alesina, Devleeshauwer & Easterly, et al., [2003]; Fearon [2003]).

However, there is also a long-standing parallel literature identifying this assumption as problematic. Indeed, the idea that people can hold multiple ethnic (or national) identities simultaneously literally goes back centuries: it was perhaps first explicitly expressed by the English historian William of Malmesbury in the twelfth century, who wrote about the existence of Northumbrian, Mercian, East Anglian and Kentish gentes (ethnic groups or nations), which were subsumed in the gens Anglorum (Bartlett, 2001, p. 43). In early modern Italy it was commonplace for the word ‘nation’ to refer to both one’s region (Piedmont, Sicily, etc.) as well as modern nations like Italy or Germany (Woolf, 2005, p. 299). More recently Max Weber distinguished between three concepts of ethnicity, namely Völkerschaft (clan), Stamm (tribe) and Volk (people), ‘each of which is ordinarily used in the sense of an ethnic subdivision of the following one’ (M. Weber, 1978 [1922], p. 392). And from the 1960s numerous anthropologists have criticized this singular, fixed conception of ethnicity, instead insisting on multiple identities that co-exist and achieve different levels of salience depending on the context (Gluckman, 1960; Gulliver, 1969). As such there is now a large literature in which it is widely acknowledged that individuals can have a multitude of identities, many of which can be considered as based on a belief in common descent and thus ethnic in nature.5 These identities can be held at the same time because they are nested within each other, such that one could conceptualize the smallest group as those with whom one shares a common (mythical) ancestor in a relatively recent generation, with the ancestor increasingly further back in the past as one moves to larger groups.6 In this context several scholars have previously discussed

5 There is also a literature which has noted a multiplicity of social identities available to people, only one of which was ethnic (Eriksen, 1993, p. 30; Okamura, 1981), or concentric versions of the same ethnic identity (Galaty, 1982).

6 To repeat, nothing I say here should be taken to mean that such ancestors ever existed in the first place, but merely that people believe in the existence of common ancestry along these lines. There is, moreover, plenty of evidence that the group of common ancestors a given set of people could share is potentially very wide, considering ever-shifting ethnic descent myths (A. D. Smith, 1998, p. 149).
instances where people choose between two ethnic identities, such that people ‘put on the ethnic hat’ that best befits a given situation (Gil-White, 1999, p. 807). To take one example from Africa, Posner (2005) argues that Zambians hold both larger linguistic identities (Bemba, Lozi, Tonga, etc.) and smaller tribal identities (Lamba, Lunda, Luvale, etc.), and choose to emphasize one or the other according to which one was more likely to benefit them at a given point of time.

However, there is no reason to stop at only two ethnic identities. In Uganda and many other parts of Africa, for instance, one could be a member of a sub-clan (for instance, the Busito sub-clan within the Nte clan of Buganda), a clan (Nte), a tribe (Buganda), a super-tribe, linguistic or regional group (Bantu-speakers or Southerners, groups which largely overlap) and a race (African). Or, for Native Americans in the United States, one could be a member of the Upper Brulé thiyospaye (band or sub-clan), Brulé clan, Lakota tribe, Sioux nation and Native American race (J. R. Hanson, 1997, p. 203). Finally, for another example from the USA, a black woman in New York could potentially identity as Trinidadian, West Indian or Black (Chandra & Wilkinson, 2008, p. 521). Figure 1.1 makes this series of groupings graphically clear.

Thus, as Gulliver (1969, p. 21) puts it, ‘at any one level the [ethnic] groupings are, on the one hand, amalgamations of those at a smaller-scale level and, on the other hand, they are constituents of the larger groupings at a wider-scale level’. Rather than thinking in terms of ‘hats from the ethnic rack’ (Gil-White, 1999, p. 807), a better way to conceptualize these different identities is as different sized Russian (Matryoshka) dolls, with individuals possessing multiple identities based on different sized groups. One could also think of a cross-section of these dolls, which would resemble a series of concentric circles representing different social groups. Indeed, this is the model presented by the noted twentieth-century psychologist Gordon Allport in his model of in-groups, moving from the smallest circle, the family, to the nation and eventually all of mankind (Allport, 1954, p. 43). A process of ethnic homogenization would thus consist of shifting one’s primary, most salient identity from a smaller doll or circle to a larger one, thereby emphasizing an identity which is shared by more people than the previous identity.

Inasmuch as these identities operate at different levels they do not contradict each other, unlike, say, attempting to be both a Muganda (member of the Buganda tribe) and a Mukiga (from Kigezi in western Uganda), or members

---

7 For other such examples of multi-level ethnic identities see Moerman (1965, p. 1224), Rousseau (1975, p. 44) and Wimmer (2008, p. 977) using case studies from Thailand, Malaysia and the USA, respectively.

8 For other similar conceptualizations of ethnicity as a series of nested identities also see G. C. Bentley (1987, p. 35), R. Cohen (1978, p. 387) and Emberling (1997, p. 307).

9 The analogy is not new; cf. Taras’s (1993) description of ‘Matryoshka nationalism’ in the former Soviet Union.
Concentric loyalties need not clash. To be devoted to a large circle does not imply the destruction of one’s attachment to a smaller circle. The loyalties that clash are almost invariably those of identical scope … A traitor who serves two nations (one nominally and one actually) is mentally a mess and socially a felon. Few people can acknowledge more than one alma mater, one religion, or one fraternity. On the other hand, a world-federalist can be a devoted family man, an ardent alumnus and a sincere patriot. (Emphases in original; [Allport, 1954, p. 44]).

It is notable that Allport claims that ‘the happy condition [by which] narrow circles can … be supplemented by larger circles of loyalty … is not often achieved’ (Allport, 1954, p. 46). My argument in this book is that such shifts happen more often than Allport, or others, have thought.
The second stage of my argument is that:

2. People choose their primary ethnic identity based on the benefits it brings to them.

To put this in the language of rational-choice scholarship, people choose which of their multiple identities will maximize their utility, or which identity ‘serves them best’ (Posner, 2005, p. 2). As with the first step this is not a particularly radical assumption, inasmuch as it is the basis for the current scholarship on identity formation in the social sciences labelled as ‘instrumentalist’. Here ethnicity is an instrument, or a means to an end, rather than an end in itself, in contrast to the ‘primordialist’ views noted above; similarly, one can conceive of ethnicity in this sense as ‘constructed’ rather than fixed, thereby generating a ‘constructivist’ understanding of ethnicity.11

That economic benefits form an important incentive in how one chooses an ethnic identity should be self-evident, since maximizing one’s economic utility in general leads to a focus on income by definition. Yet surprisingly this step has remained slightly hidden or obscured in the literature on ethnicity, in part because much of the literature on ethnic change and identity has not come from economics but instead from other disciplines. In particular many instrumentalist scholars have focussed on the political motivations for choosing ethnic identities, such that it is often assumed that ‘the sole raison d’être of ethnicity and ethnic organization lies in its political functioning’ (Eriksen, 1993, p. 55). Thus for Posner (2005), for instance, the focus is on how political institutions (specifically the nature of the party system) alter ethnic identities, while for Laitin (1998) and Posner (2004) the focus instead is on the construction of state borders. Part of the problem is that in none of these cases was the research focus on explaining ethnic change per se, as opposed to demonstrating whether and how certain political phenomena could alter ethnic identities.

Here, however, I focus directly on ethnic change as a research topic, which leads me to focus on the role of income and economics instead of only on politics. More specifically, I ask whether broad trends in economic transformation can explain ethnic change. Ethnic change can be conceptualized qualitatively, as already discussed, but can also be measured quantitatively using measures of the ethno-linguistic fractionalization (ELF) index across time. The ELF index for a given state or area is calculated using the Herfindahl concentration formula, namely by summing the squares of the percentages of all ethnic groups larger than 1 per cent of the population and subtracting this sum from one. Thus an ELF index for a very homogenous country like Haiti, Portugal or South Korea would be a number close to zero, while one with a high amount of ethnic diversity like India, Liberia or Uganda would have a

---

11 Some authors identify constructivism and instrumentalism as different strands of thought in the study of ethnicity; I prefer to follow Henry Hale (2004) in thinking of them both as a single programme.
score close to one. Similarly, a decreasing ELF index would indicate a process of homogenization, while a rising ELF index would indicate increasing levels of diversity. (I discuss the ELF index in more detail in Chapter 2.)

I now move to my third step, namely that:

3. The most important factor explaining ethnic change – specifically assimilation – in the modern world is industrialization.

I argue here that industrialization encourages assimilation as it alters the economic incentives of holding both smaller rural ethnic identity and broader urban ethnic identities. In the first instance industrialization lowers the value of emphasizing rural identities as it lowers the value of rural land relative to other parts of the economy. In the second instance industrialization raises the value of broader industrialized or urban identities which can aid urban residents in getting jobs in the modern economy and/or getting support from other rural–urban migrants in the difficult urban landscape. I specify both of these processes in more detail below.

To understand the effect of industrialization on rural identities it is necessary first to examine the nature of ethnic identity and ethnic diversity in a broad historical context. Recently many economists, ecologists and anthropologists have used quantitative methods to show that latitude is inversely correlated with ethno-linguistic diversity (Ahlerup & Olsson, 2012; Cashdan, 2001; Collard & Foley, 2002; Green, 2013; Sutherland, 2003), such that countries closer to the equator have higher levels of ethnic diversity. Other researchers have similarly found a positive correlation between ethnic diversity and elevation (Cashdan, 2001; Sutherland, 2003; Wimmer, 2015) and differential land endowments (Michalopoulos, 2012).

There are two possible interrelated reasons for these relationships. First, warm tropical environments with predictable climates (as regards the variability of temperature and rainfall) are ideal for becoming self-sufficient in food and thereby create few incentives for inhabitants to migrate or form social ties with people across a large amount of territory (Nettle, 1996). The same logic applies to differential land endowments, which lead human groups to become specialized in growing certain types of crops and thus discourage migration to other areas that are not suitable for the same crops (Michalopoulos, 2012). Second, even if humans had wanted to migrate, mountains, different disease environments and dense tropical rainforests can create physical barriers and disincentives to movement and thereby lead to cultural isolation. In both senses humans are no different from other species, inasmuch as there is also an inverse correlation between latitude and species diversity (Cashdan, 2001). Indeed, the first person to note this relationship was Charles Darwin in his trip to the Galapagos Islands, where he found twenty-five different species of finches which, he hypothesized, had resulted from the effects of being isolated on separate islands.

This quantitative literature on the role of isolation in promoting ethnic diversity has a parallel in qualitative social science as well. From the
Asia-Pacific region there is evidence that poor transport links and rugged terrain have promoted high levels of ethnic or cultural diversity in such contexts as Sarawak (Malaysian Borneo) and Papua New Guinea (Reilly, 2001, p. 174; Rousseau, 1975, p. 34). In pre-colonial Africa ‘ethnicities were often founded on the special skills needed to exploit different ecological niches’ such that the Fulani of West Africa became associated with cattle-herding while in East Africa the Ogiek and Kikuyu became known as hunters and farmers, respectively (Lonsdale, 2012, p. 34). Finally, in South-East Asia Scott (2009, pp. 259–61) has suggested that ethnic groups and their boundaries have formed around particular livelihoods. Indeed, he quotes one European visitor to China from the 1870s who noted that ‘the frontiers between different types of soil, between farming and herding, and between Chinese and Mongols coincided exactly’ (quoted in Scott [2009, p. 262]). In regions with similar ecologies, however, it is more common to observe ‘the homogenizing effects of a common agrarian regime,’ as with the Shan of South-East Asia where being Shan became associated with padi planting (Scott, 2009, p. 252). In other words, as with Michalopoulos (2012)’s argument it is ‘rugged topography and relative isolation’ which leads to ethnic diversity for Scott (2009, p. 265).

Thus in the pre-modern or pre-industrial world isolation encouraged greater ethnic diversity. In such isolated ‘folk societies’ (Redfield, 1947) it would have made no sense to identify with a foreign ethnic group inasmuch as livelihoods and income within the group depended on other members of that group. In contrast, however, the incentives for ethnic identification change with industrialization as people leave the rural agrarian economy to move to the industrial urban sector, where access to rural land, ethnic-specific agricultural skills and support from one’s rural community are all relatively unimportant in making a living. Instead, what is valuable to the new industrial working class is their ability to earn an income through their labour, which can be enhanced by shedding their previous narrow rural identity in favour of a larger identity shared by more people. In other words, as both people and profits move from rural to urban areas, access to labour become more important than land in the acquisition of income and wealth.

Theoretically the first and most forceful such statement identifying the homogenizing incentives of industrialization comes from Karl Marx and Friedrich Engels in their early writings from the 1840s and 1850s. In works such as the German Ideology (1846) and the Communist Manifesto (1848) Marx and Engels focussed their attention on the division between ‘town and country’ (Marx & Engels, 1947 [1846], pp. 68–9). More specifically, they claimed that in the countryside individuals are ‘united by some bond: family, tribe, the land itself, etc.’ and the economy is guided by an exchange between people and the natural environment rather than through other people. The ownership of land is the basis for the economy and thus people remain attached to the land, with ‘their mode of production isolating them from one another, instead of bringing them into mutual discourse’ (Marx, 1978 [1852], p. 608).
Individual peasants may be said to be members of an ethnic group, class or nation, but without any ‘community [or] national bond’ they are merely like a ‘sack of potatoes’ in Marx’s memorable phrase.

However, upon entering the scene, the bourgeoisie ‘create enormous cities, greatly increase the urban population as compared with the rural, and thus rescue a considerable part of the population from the idiocy of rural life’ (Marx & Engels, 1978 [1848], p. 477). Modern industrialization sees a huge shift in the basis of economic production as the basis for the economy and economic growth shifts from land to labour. Here members from previously isolated and separate rural communities come together to work, thereby making manifest for the first time the division between the ‘two great classes’ of the bourgeoisie and the proletariat. While in rural areas labourers held location-specific capital, in the industrial sector they lose their distinctiveness and become part of a ‘floating’ army of workers, whereby their exploitation at the hands of the bourgeoisie creates the impetus for class consciousness and subsequent revolution.

Thus, while not claiming explicitly that industrialization promotes assimilation, Marx and Engels were nonetheless clear about the way in which structural transformation created incentives for social homogenization, in two ways. First, the bourgeoisie needed to create a homogenous workforce – or ‘privates in the industrial army’ (Marx & Engels, 1978 [1848], p. 479) – that it could employ in its factories and shops, a process which I henceforth label the ‘top-down’ mechanism inasmuch as the process of ethnic homogenization is largely a result of the efforts of the ruling elite. However, Marx is explicit that the process of class formation is a byproduct of modern capitalism rather than a deliberate outcome directed by the bourgeoisie, which has no incentive to create a large, united class opposed to its interests. Instead, there is a second process of homogenization at work whereby members of the proletariat have strong incentives to co-operate through unions and political parties to fight the exploitation of the bourgeoisie. Indeed, Marx and Engels’ famous call at the end of the Communist Manifesto for workers of the world to unite is essentially based on their belief that, despite their varied backgrounds and cultural/ethnic differences, the advent of the modern industrial economy means that workers’ interests now lie together with each other. This second process I label the ‘bottom-up’ mechanism, since here homogenization is the result of the actions of the non-elite based on what they perceive to be their best interests.

12 Marx saw the parochial and isolated nature of rural life as universal: in his writings on India, for instance, Marx writes of the ‘idyllic village-communities’ where life is ‘undignified, stagnant, and vegetative’ (Marx, 1978 [1853], p. 658).

13 Marx and Engels clearly distinguish the economic structure of pre-industrial cities from modern industrial cities: in the former case the division of labour was between different guilds while in the latter case it is instead between classes within the same workshop (Marx & Engels, 1978 [1848], p. 474).