Chapter 2
Ethnicity, Language and Populations

Abstract In an increasingly globalised world, ethnicity, or the question of defining one's personal identity with reference to a "group of common descent" has become very prominent in political as well as scientific debates. Ethnicity is socially constructed and defined subjectively by a combination of aspects related to a group's ancestry, cultural customs, language, religion, national identity, kinship networks and even physical appearance. This slippery nature has made ethnicity the most difficult variable to conceptualise and measure in social as well as medical and biological research. This chapter reviews the main definitions of ethnicity, and the related concept of "race", as well as some of the approaches to measure it. It then proposes a multidimensional approach that conceives ethnicity as an outcome, disentangling some of the processes that end up constituting a group's identity. Finally, it justifies the need for alternative measures of ethnicity, one of them being the use of personal names' origins and forename-surname networks in attempting to disentangle such processes.

In the last decade and a half, there has been an explosion of interest in issues of ethnicity, nationalism, race and religion, around a renewed preoccupation with the question of defining and asserting collective identities. This trend has contradicted the prediction made in the 1920s by Max Weber (1980 [1921]) who stated that "primordial phenomena" such as ethnicity and nationalism would decline in importance and eventually vanish as a result of modernisation, industrialisation and individualism. On the contrary, the change of Millennium has brought opposition between an ever-expanding globalisation and an upsurge in identity, an antagonism that is key to understanding the way that our world and our lives are being shaped (Castells 1997). Collective identities are formed and expressed as a resistance movement to cultural homogenisation (Castells 1997) in a struggle for political power in multicultural societies (Kertzer and Arel 2002). This is set in a context of a diminishing role of the nation-state, with its political power being devolved to the regions and cities as well as taken away by international institutions and a new global order. Combined with these trends, long-established nineteenth century

national identities are being eroded in an era of migration, characterised by the increased intensity and complexity of its flows (Castles and Miller 2003).

In such circumstances, governments and social scientists have struggled to keep track of the reality of a rapidly changing population that is constantly re-defining its collective identities (Skerry 2000). Although highly contested, the practice of classifying the population into discrete groups according to race, ethnicity or religion has made a strong re-appearance in many countries' national censuses (Howard and Hopkins 2005; Kertzer and Arel 2002; Mateos 2014; Nobles 2000). Such questions in the censuses not only quantify the size and geographical extent of collectively pre-perceived racial, ethnic and religious groups, but more interestingly help to reinforce the self-identity of those groups or accelerate the emergence of new identities (Christopher 2002) by solidifying transient labels (Howard and Hopkins 2005).

Because of the subjective nature of collective identities, the categorization process (the problematic definition of ethnic groups' boundaries and labels) has been a significant issue in social science (Peach 1999). Following an impassionate debate around the essentialism of ethnicity labels (Modood 2005), there seems to be a consensus, at least in population studies, that the classification of the population into ethnic groups has proved useful to fight discrimination and entrenched health and social inequalities (Bhopal 2004; Mitchell et al. 2000). There is a vast literature that demonstrates the persistence of stark inequalities between ethnic groups, especially in terms of health outcomes, access to housing and labour markets, educational outcomes and socioeconomic status (Frazier et al. 2003; Mason 2003). As long as such inequalities between population subgroups persist, no matter how these are defined or perceived, the use of ethnic group definitions and labels will be useful to denounce them and fight against their causes. However, many of the current ethnicity classification practices have proved very inappropriate to uncover the true nature of specific factors of inequalities.

This first chapter sets out the general context of the different intersections between the ontology of ethnicity and its measurement in social science. Section 2.1 directly addresses the ontological issues behind the concepts of ethnicity and its multidimensional characteristics, taking a broad perspective drawn from the anthropological, sociological, geographical and health literatures. Furthermore, Sect. 2.2 complements the ground laid down in the previous section with an extensive review of the different ways in which ethnicity is measured in different contexts, identifies the key issues of measurement and investigates how they affect the analysis of ethnicity.

2.1 Constructs of Race and Ethnicity

The study of ethnicity and race in multicultural societies and cities is probably one the most problematic phenomenon that social scientists face today. Ethnicity and race are very controversial variables in scientific inquiry, and during over 150 years

of speculation, biologists, anthropologists and geneticists have demonstrated over and over again that these terms are both socially constructed and lack any biological reality (Cavalli-Sforza 1997). Ethnicity relates to a person's inner sense of collective identity, and its definition requires contact between differently perceived groups to create a difference. Such contact has exponentially increased in the last decades as populations, cities and neighbourhoods have become increasingly multiculturally diverse and globally connected (Castles and Miller 2003). If "national identity requires a collective work of amnesia" Renan (1990 [1882]: 11), it could be argued that in today's context of globalisation and erosion of nineteenth century nation-state identities, ethnic identity requires a collective work of "remembrance and nostalgia".

The definition of ethnicity and race are controversial because identification is subjective, multi-faceted and changing in nature and because there is not a clear consensus on what constitutes an "ethnic or racial group" (Coleman and Salt 1996; Office for National Statistics 2003). Moreover, ethnicity and race classifications have become a key factor of political power in the growing arena of identity politics (Skerry 2000). The power struggle between competing collective identities for institutional recognition through official ethnicity classifications is especially manifested at local level, where such recognition brings resources and solutions for locally perceived problems, financial aid, political representation, and benefits associated with positive discrimination initiatives (Kertzer and Arel 2002).

However, the main purpose for which race and ethnicity started to be officially classified and measured in national statistics in a number of developed countries in the last decades bore little correspondence with this identity politics struggle. It directly emanated from the need to monitor progress in equality legislation, introduced to prevent racial discrimination and reduce ethnic inequalities after the 1960s (Peach 2000), in particular the American Civil Rights Movement (1954–1968) and the UK Race Relations Act (HM Government 1976). Such legislation and the population classifications derived from them were only concerned with people seen of "darker skin colour", following non-European post-war migration to America and Europe and the deeply rooted black discrimination in the US (Coleman and Salt 1996). Such perception of difference evolved into the term "visual minorities", commonly used in Canada and other countries, which goes beyond differences in phenotype and encompasses any other visual element of cultural difference such as religious symbols in clothing or hairstyle. Today, most countries' collective identity classifications go beyond visual or biologically rooted concepts, such as "race", and use the broader cultural term of ethnicity. Such issues of definition of race and ethnicity and their criticisms are discussed in the next paragraphs within this section.

2.1.1 Race

The history of how, during the age of European colonialism, scientists identified races and ranked them according to their biological and social value, with the "White-European race" always ranking on top, is unfortunately well known (Gould 1984). They justified such rankings based on claims of intelligence hierarchies using measurements of the size and shape of the head, and even the contents of the brain (Gould 1984), with the underlying value that biology determined social position; in short, biological determinism (Bhopal 1997). This type of research whereby human populations were divided into sub-species, mainly on the basis of visible physical characteristics, was used to justify slavery, imperialism, anti-immigration policy, and the social status quo (Bhopal 1997). These views are well depicted by the illustration shown in Fig. 2.1, taken from a 1968 US primary school Atlas.

This view of human races as discrete biological constructs, was dominant for most of the nineteenth century and beyond until its abandonment with the defeat of the Nazis at the end of the Second World War (Bhopal 2004). Attached to the ideas of the Nazis were the Eugenic theories, which sought the improvement of the "human race", in particular the "Aryan race". A book titled "Outline of Human Genetics and Racial Hygiene" was published in 1921 by the geneticist Fritz Lenz, a leading advocate of the Aryan ideology, and is claimed to have been very influential in Hitler's (1925) own book "Mein Kampf", where he set out his political beliefs about German racial superiority (Olson 2002).

Today race is defined in the Cambridge English Dictionary as "a group, especially of people, with particular similar physical characteristics, who are considered as belonging to the same type, or the fact of belonging to such a group" (University of Cambridge 2004). Therefore, it is a subjective "consideration of belonging" that makes it a social construct. There is a general agreement, forged through the last four decades of population genetics research, that the concept of race is socially constructed, and cannot be explained by genetic differences between human groups (Cavalli-Sforza 1997). But even though none of the numerous "scientific" racial classifications has stood the test of time (Bhopal 2004), current "race" classifications remain influenced by "biologically rooted" racial stereotypes (Graves 2002). Consequently the concept of "race" is still strongly used in many countries, such as the U.S., when subdividing populations according to their ancestral origins. The persistence today in the U.S. of the concept of race, and hence the use of racial classifications in administrative records and academic studies, may be traced to the legacy of the American Civil Rights Movement (1954–1968) and the legislation subsequently introduced to prevent racial discrimination. Although the contemporary concept of "race" has partially lost its roots in distinguishing differences in physical appearance alone (phenotypes), it is still loaded with ideological assumptions about innate, hereditary, ranked differences between groups of people (Chapman and Berggren 2005).

OF THE PEOPLE WHO INHABIT THE EARTH.

How many people are there in the world?

There are upwards of a billion—(1,000,000,000.)

How are the people in the world divided?

The people in the world are divided into five principal races, named according to their color and residence.

Name the five races in the world.

The five races are the White or Caucasian; the Yellow or Mongolian; the Black or African; the Brown or Malay; and the Red or American.

What is known of the White race?

The White race is superior to the others, and is found in Europe and America.

THE WHITE RACE.

Of the Yellow race?

The Yellow race is found in Asia; the best specimens are in China and Japan.

Of the Black race?

The Black race is found in Africa, and is commonly called the Negro race.

Of the Brown race?

The Brown race inhabits the islands of the Pacific Ocean.

THE YELLOW RACE.
(A Chinese Laborer.)

THE BLACK RACE.
(An African Chief.)

Of the Red race?

The Red race includes the Indians of North and South America.

THE BROWN RACE.
(A New Zealand Chief.)

THE RED RACE.
(An Indian Chief.)

Fig. 2.1 Nineteenth century illustrations of 'the races of the world'. *Source*: (Mitchell 1868:11)

However, the debate surrounding biological differences of human groups has not been closed, but actually moved on to a new stage in the era of individual human genetics. In a special issue of *Science* magazine, published in 2005 to commemorate its 125th anniversary, two of the "125 big questions that face scientific inquiry over the next quarter-century" (Science 2005: 5) are very closely related to this debate:

"What are human races, and how did they develop? Anthropologists have long argued that race lacks biological reality. But our genetic makeup does vary with geographic origin and as such raises political and ethical as well as scientific questions" (Science 2005: 100. Emphasis added)

"To What Extent Are Genetic Variation and Personal Health Linked?" (Couzin 2005: 85)

There is a growing belief, in the health and anthropological literature, that the biological concept of race made a strong come back at the turn of the Millennium, hand in hand with the genetics revolution in science (Kahn 2005). In this era of race genetics and genetic medicine (Nature Genetics 2001), "Gene hunting [has become] the new research colonialism" (Pearce et al. 2004: 1071), in which scientists try to identify key differences in gene frequencies between different "populations". The key to mapping DNA groups therefore lies in the definition of such "populations", which are again socially constructed based on geographical, anthropological and historical assumptions (M'charek 2005).

In order to overcome the biological determinism implicit in the term "race", and to include other non-biological factors that make us perceive human groups as different from each other, the concept of "race" has been rapidly abandoned in scholarship as well as government in favour of that of "ethnicity". This trend has been observed in the last three decades of the twentieth century, primarily outside the US (Oppenheimer 2001), and is especially well documented in the health literature (Afshari and Bhopal 2002). However, this trend is not without problems, since it assumes that both terms can be used interchangeably, as if it both described the same quality—despite this assumption being disproved by many authors (Bhopal 2004).

2.1.2 Ethnicity

The word ethnicity derives from the Greek word ethnos, meaning a nation, and the term "ethnic group" is considered to have been introduced by Max Weber in 1921. He defined ethnic groups as "[t]hose human groups that entertain a subjective belief in their common descent because of similarities of physical type or of customs or both, or because of memories of colonization and migration (. . .) it does not matter whether or not an objective blood relationship exists" (Weber 1980 [1921]). Therefore, at the core of the concept of ethnicity is a subjective belief of common origins without the necessary existence of genetic linkages or physical similarity. This concept is thus closely linked to the question of an individual's identity, which is defined by the characteristics of the ethnic group to which he or she recognises belonging. Amongst the main reasons for such perception of self-identity are certain shared characteristics, including physical appearance, but most importantly geographical and ancestral origins, cultural traditions, religion and language (Bhopal 2004). This brings to the fore the multi-dimensionality of the notion of ethnicity. Bulmer proposes one of the most widely accepted definitions of "an

ethnic group [as] a collectivity within a larger population having real or putative common ancestry, memories of a shared past, and a cultural focus upon one or more symbolic elements which define the groups' identity, such as kinship, religion, language, shared territory, nationality or physical appearance" (Bulmer 1996: 35). Understood as such, ethnicity is considered to differ from race, nationality, religion, and migrant status, sometimes in very subtle ways, although it is considered to include traits of these other concepts as well (Bhopal 2004).

Therefore, at the core of the concept of ethnicity is the question of an individual's identity, which is defined by the characteristics of the ethnic group that he or she considers herself to belong to, always understood in a contextual rather than in an essentialist way (Peach 1996: who himself might be considered Welsh in England, British in Germany, European in Thailand, and White in Africa). The social context in which the ethnic group is defined is therefore key to understanding its identity. This idea stems from one of the more interesting facts observed during the processes of ethnic group formation; not only a firm belief in group affinity is required for group identities to emerge, but this is usually defined in opposition to other groups perceived as being culturally different and with whom contact is required (Eriksen 2002). In other words, if there is no contact with other groups that are perceived as "culturally different", the identity of an ethnic group does not emerge. For example, the concept of a Hispanic ethnic group only emerged in the US during the 1960s and 1970s, when large numbers of Spanish speaking immigrants from many countries and their descendents found a common identity through a shared language and migration history in an English-speaking country. Not only did the "host culture" consider them as one group, but Spanish-speakers from Latin America considered that this new identity would make them stand out in the US in a much stronger way than with their individual national identities (Skerry 2000). The paradox is that no Spanish speaker outside the US would consider himself or herself as "Hispanic", and the group's homogeneity is difficult to sustain (Choi and Sakamoto 2005). This important appreciation of contact between differently perceived groups explains why the debate on ethnic identity has grown since the end of the Cold War in developed countries (Castells 1997). This recent trend is explained by the disappearance of the communist-capitalist bipolar world and its political antagonism that prevented mass population movements and the redrawing of national borders, the diminishing role of the nation-state, globalisation and the growth of nationalisms, and a growing number of different human groups living amongst each other in large numbers (Castles and Miller 2003).

2.1.3   Criticisms

Nonetheless, the characteristics that together define ethnicity are not fixed or easily measured, so ethnicity is considered a subjective, contextual, transient and fluid concept (Senior and Bhopal 1994), and probably the most controversial subject of study in social science (Nobles 2000). The fluidity of the concept of ethnicity is at

the root of the anti-essentialists' critiques, who challenge the whole idea of trying to classify people into discrete and immutable categories, such as social classes but especially ethnic groups (Brubaker 2004). These authors favour the concept of "identities" which are subjective, fluid and always evolving, where people can assign themselves to several, even overlapping categories which, taken together, may better reflect the complexity of their lives (Pfeffer 1998). Even the American Sociological Association describes race (in the US research context) as "a social invention that changes as political, economic, and historical contexts change" (American Sociological Association 2002: 7). Although, as has been mentioned above, there is a consensus that the modern concept of race is not equivalent to ethnicity, the differences between the two are still widely ignored by researchers (Comstock et al. 2004). This confusion makes the understanding of the separate processes of inequalities, arising from racial, or cultural/ethnicity factors, even more difficult and controversial.

Other authors such as David Harvey (2005) relate current issues of ethnicity and race difference with more traditional structural differences in class identity;

> "*Popular as well as elite class movements make themselves, though never under conditions of their own choosing. And those conditions are full of the complexities that arise out of race, gender, and ethnic distinctions that are closely interwoven with class identities.*"
> (Harvey 2005: 202)

This contention seems to suggest a situation of "old wine in new bottles", in which new identities formed around minority groups (according to race, ethnicity, gender, sexuality, age, or disability) have replaced old divisions along social class lines in the explanation of socio-economic inequalities.

Going back to the concept of ethnicity, because it is considered a core element of personal identity, the current preferred method for ascribing one's ethnicity in research and government statistics is self-assessment. However, since the categorizations of ethnic groups are usually pre-classified and individual choice is constrained to choosing amongst them, the concepts of ethnic groups themselves are also considered an externally imposed identity (Senior and Bhopal 1994). Therefore, the definition and measurement aspects of identity are closely related and cannot be studied in isolation. The problem of ethnicity measurement is dealt with in the next section. Ethnicity, rather than the more biologically rooted concept of race commonly used in the US, will be used from now on in this book. Ethnicity is the concept most widely used to identity population groups that share an ancestral and cultural origin, and thus a much closer term to the main theme of this book, hence its prominence in this book's title.

2.2   Measurements of Ethnicity

Following from the complex definition of ethnicity presented in the previous section this section will review the issues around the difficult task of measuring ethnicity in order to classify people into ethnic groups.

2.2.1   Measurement in Official Ethnicity Classifications

It should be obvious by now why the measurement of ethnicity is problematic; because ethnic identification is subjective, multi-faceted and changing in nature and because there is not a clear consensus on what constitutes an "ethnic group" (Coleman and Salt 1996; Office for National Statistics 2003). However, as has been justified in the introduction of this chapter, the measurement of ethnicity is today useful for a wide range of purposes in many countries, especially to reduce ethnic inequalities and to understand our recent past. This puts pressure upon government statisticians who try to cope with surges of interest in collective identity formation and with the struggle of States to monitor and sometimes try to shape these processes (Kertzer and Arel 2002). Even when a consensus in social statistics is reached, with time the action of statisticians cannot be detached from their consequences on the reality being measured, and as Barrier (1981) puts it; "The census imposes order of a statistical nature. In time the creation of a new ordering of society by the census will act to reshape that which the census sought to merely describe" (Barrier 1981: 75).

The national Census of Population comprises the major classificatory effort of a society, and has been described as a sort of communal "family photograph" that is only taken every 10 years (Skerry 2000). Therefore, the social processes and groups that appear in such photograph are of high importance, since census enumeration brings with it political and economic power (through representation and funding). As such, the classification of the population into the groups of common ancestry used in the census brings with it a halo of official statistical recognition that transcends the census enumeration exercise itself and determines all sorts of possibilities for an ethnic group during decennial inter-censal periods and beyond (Skerry 2000). As such, in most countries the de facto "gold standard" for ethnicity measurement usually emanates from the categories created by the national population censuses (Kertzer and Arel 2002).

The UK Office for National Statistics (ONS) recognises that measurement of ethnicity should be done in a way that is sound, sensitive, relevant, useful, and consistent over some period of time (Office for National Statistics 2003). However laudable these statisticians' principles, Skerry (2000) depicts very well the tension in the US Census Bureau "between the extremely technical character of the census and the emotional, highly symbolic nature of race politics" (Skerry 2000: 4). These types of frictions were behind the reasons why, despite having been considered

since 1971, an ethnicity question was not introduced in the UK until the 1991
Census (Coleman and Salt 1996), why it is still not asked in many countries (for
example in France or Spain), and why it has created so much controversy before
and after each US census during the last decades (Nobles 2000). An early quote
from the introduction in the US of an official racial and ethnicity classification
summarises well this point:

> "*These classifications [set in the Racial and Ethnic Standards for Federal Statistics and
> Administrative Reporting] should not be interpreted as being scientific or anthropological
> in nature*" (Office for Management and Budget 1978: 19269)

Even when national consensus is reached, a further problem arises when trying to
perform international comparisons between national censuses, since the terms used
to describe ethnic groups are developed within each country in response to their
own particular historical processes of ethnogenesis (Aspinall 2005). In the round of
population censuses conducted in 2000/2001, 141 countries collected information
about the ancestries or identities of their populations, using questions on one or
more of the following dimensions of identity; ethnicity, race, indigenous/tribal
origin, and nationality (Morning 2008). At the time of writing, a similar evaluation
exercise has not been conducted for the 2010/2011 round of censuses. In a thorough
comparison of 20 countries 2010–2011 census questionnaires (Mateos 2014) we
have identified 27 different questions on various identity aspects that could be
summarised under six major themes: (1) Residency and migration, (2) Citizenship,
(3) Country of birth, (4) Ethnicity/Race/Ancestry, (5) Language, and (6) Religion.
However, international comparisons are highly limited because of the different
ontologies of ancestral origin and identity that underlie each of the classifications. A
detailed study of such national classifications would entail much more space than
available in this chapter, and hence it will mainly focus on the United Kingdom
experience, arguably the European country with a longer tradition in measuring
ethnicity.

2.2.2   Issues with Official Ethnicity Classifications

Despite their widespread influence, there are three major problems with the way
ethnicity is currently officially measured in most developed countries. First, eth-
nicity is usually measured as a single variable, that of an "ethnic group" into which
the individual self-assigns his or herself from a classification of a reduced number
of classes, with no leeway to represent any characteristics of the multi-faceted
nature of self-identity described above. This problem has been partially addressed
in some Censuses, such us in the US, in which respondents were able to choose
from more than one "race/ethnic group", although it has created a new issue of
comparability across time and between different combinations.

A second problem is that pre-set ethnic classifications are used as opposed to just an open question, in which the responses are then arranged according to the most meaningful common identities. This is of course justified with the need to facilitate the creation and comparison of the resulting statistics over time and between different information sources (Office for National Statistics 2003). However, as mentioned before, these categories have proved not to reflect the complex heterogeneity found within each group (Agyemang et al. 2005; Connolly and Gardener 2005; Rankin and Bhopal 1999).

A third problem arises from the method of determining ethnicity by self-assessment, which comprises the current consensus across datasets and the literature (Bhopal 2004), as opposed to it being assigned by a third person or a computer according to some established measurable criteria. As a result of self-classification, the ethnicity of the same person can vary through time, since perceptions of individual and social identity changes over time (Aspinall 2000) and are influenced by the type of ethnicity question asked (Arday et al. 2000), the definitions of categories offered (Olson 2002), and the country and method of data collection. Although this is not the aspect of ethnicity classification that is the most highly debated, self-defined ethnicity has been deemed as "unhelpful" (McAuley et al. 1996).

In addition to these three major issues of official ethnicity classifications, an additional recognised problem is the lack of routine collection of ethnicity data in most government or public service datasets.

2.2.3   The Limits to Comparability Between Research Studies

Even when ethnicity information is collected, its consistency and comparability is usually very poor. As a consequence, research on ethnicity has been hampered by a lack of common methodologies in the collection and treatment of ethnicity information (Whitehead 1992). Different studies define ethnicity in different ways, and create independent classifications and non-comparable methods of data collection (Choi and Sakamoto 2005). Decisions taken in this respect are only based on the tactical considerations deemed most appropriate for each context, while making explicit neither the methodology nor the classification. The inevitable consequence is that results cannot be correctly interpreted and compared between studies.

The problem of lack of comparability is especially critical in research about differential outcomes by ethnic group. Comstock et al. (2004) summarise very well the extent of this problem in public health research. They conducted a comprehensive review of 1,198 articles published in the *American Journal of Epidemiology* and the *American Journal of Public Health* from 1996 to 1999, and found 219 different terms to describe just eight core "ethnic groups". The detailed descriptions given by epidemiologists to these ethnic groups are worth exploring; hence they are included here in Table 2.1. Moreover, the authors denounce the frequent failure of researchers to explicitly define the ethnic categorizations and their context of use, to

Table 2.1 Terms used to refer to ethnicity and race in articles published in two American public health journals: 1996–1999

Asian (n = 37)	Black (n = 16)	Hispanic (n = 46)	White (n = 32)	Other (n = 38)	Unknown or missing (n = 9)	Mixed race or ethnicity (n = 7)	Additional terms (n = 34)
Asian	African	Caribbean Hispanic	Anglo American	All others	Missing	Biethnic	American
Asian/Oriental	African American	Central/South American	British Whites	Missing or other	Not indicated	Biracial	Ashkenazi Jews
Asian/Pacific Islander	Black Americans	Cuban	Caucasian	Multiple/other/ unknown	Not ascertained	Black/White	Canadian
Asian American	Black and other	Cuban American	Caucasian/other	Neither Black nor White	Not classified	Mixed ethnicity	East Indian
Asian or Filipino	Black Hispanic	Dominican	European	Non-Black (all others)	Unknown/ refused	Multiethnic	Egyptian
Asian or other races	Black, non-Hispanic	Foreign born Latina/Latino	Non-Hispanic White	Non-American Indian		Multiracial	Eskimo/Aleut/ Alaskan
Asian or Pacific Islander	Black, other	Hispanic	Scandinavian	Non-Anglo cultural groups		Partly native American	Foreign born
Asian other	Black/other races	Hispanic/Puerto Rican	Southern European	Non-Latino			Hawaiian
Asian-Indians	Blacks and other minorities	Hispanic = White, Spanish surnamed	White/Anglo	Nonminority (other)			Jewish
Cambodian	United States/ Blacks	Hispanic all races	White/other	Non-US born			Middle Easterns
Chinese		Hispanic American	White American	Non-White/ Hispanics			Native American
Chinese (Americans)		Hispanic Black	White and Asian	Non-White ethnic minority			Pakistanis
Filipino		Hispanic ethnicity	White and other	Non-White or Hispanic			United States born
Japanese		Hispanic origin	White ethnicity	Other/mixed			

	Hispanic surname	White Hispanic	Other minorities
Japanese American	Hispanic White	White non-Hispanic	Other races/ ethnicities
Korean	Latin American	White Spanish surnamed	Other non-Hispanic
Korean (Americans)	Latino/Hispanic		Other non-White
Laotian	Latino not Black		
Non-Hispanic Asian	Mexican		
Orientals	Mexican American		
Pacific Islander	Mexican immigrants		
Southeast Asian	Non-White Hispanic		
Thai	Other Hispanic		
Vietnamese	Puerto Rican		
	South American		
	Spanish		
	US-born Hispanic		
	White-Hispanic of Mexican ancestry		

The number in the column headings ($n=$) refers to the total number of terms found per ethnic group, of which only the most common examples are reproduced in the table

Source: Adapted from Comstock et al. (2004: 614)

differentiate between race and ethnicity, to state the study methods used, and to significantly discuss the results. Bearing in mind that the large collection of articles were drawn from just two journals of the same scientific discipline in the same country, where research on ethnic disparities has a longer tradition, this issue poses a crucial problem that requires "continued professional commitment [. . .] to ensure the scientific integrity of race and ethnicity as variables" (Comstock et al. 2004: 611). This problem has been also identified by other authors, and defined as an ontological problem that constitutes "a problem with basics" (Bhopal 2004: 441).

It is important to mention here the efforts made, especially in health research, to overcome the comparability issues in ethnicity studies. In the UK, this debate began following the 1991 Census inclusion of the ethnicity question and its mandatory recording in hospital admissions since 1994. Most of the main issues with the official ethnicity classifications described in this section have already been pointed out by Senior and Bhopal (1994), and have been highly debated during the last decade, with important contributions by Peter Aspinall (2002, 2005, 2007, 2009), Raj Bhopal (2004, 2007), and Bhopal and Donaldson (1998). These and other authors agree that researchers in health and ethnicity should use comparable ethnic classifications and make explicit the meanings of the ethnic group categories selected, the criteria use for such selection, their method of ascribing ethnicity to individuals, and give precise explanations of differential health outcomes by each of the ethnic groups studied. Unfortunately, this objective is still far from becoming a reality, and even more so outside ethnicity and health research.

Taken together, the issues of lack of reflection of the multi-dimensional nature of ethnicity, the use of just a few pre-defined coarse categories, the variability of self-assignment of ethnicity, the lack of routine collection of ethnicity information, and its low quality and comparability, present major impediments for researchers and public policy decision makers. Their consequences are that researchers are prevented from measuring socioeconomic inequalities, equity of access to and uptake of public services by ethnic group, and demonstration of compliance with anti-discrimination and equal opportunities legislation, in an increasingly multicultural population.

2.2.4 Alternative Measurements of Ethnic Difference

As a consequence of the lack of ethnicity data availability, other proxies, such as country of birth, have been used to ascribe a person's ethnicity when it is not known (Marmot et al. 1984; Wild and McKeigue 1997). Despite its utility to classify migrant origins, with growing numbers of second generation migrants, the proportion of the "ethnic majority" people born abroad, and migrants born in "intermediate" countries (i.e. East African Indians that migrated to the UK), this method has become increasingly inappropriate (Harding et al. 1999). In the UK 2011 Census, less than half of the ethnic minority population was recorded as born outside the UK, and of those born abroad almost half of them hold UK passports. Furthermore,

many health and demographic studies use country of birth from death certificates, which rely on an informant and may be less accurate than the census, when the person is still alive to provide the information (Gill et al. 2005).

In some countries where the concept of "foreigners" (as opposed to nationals or citizens) is still used as a proxy for ethnic minority, such as Germany, Spain or France, the main variable used to classify populations by origin is nationality, which was only recorded by the UK Census in 2011. This proxy is also problematic since it can change over time, some people retain more than one nationality, and usually second generation migrants acquire the host country's nationality.

A third alternative method employed as a proxy for ethnicity is the analysis of personal names origins. Personal names are in principle good indicators of ethnicity, at least in relation to the immediately prior generations, that gave the forename to their descendants and probably exercised some preference in the surname. After migration to another country or region, names can probably be viewed as a kind of "self-assignment" of ethnicity that is likely to have strong links to the language, culture and geography of a person's ancestry. Names have been used in particular to identify the main ethnic minority populations in some "destination countries", with a relatively good degree of accuracy. This alternative method forms the core methodology of this book, and as such will be further reviewed in detail in Chaps. 3–6, and built upon through an innovative methodology in subsequent chapters. Therefore, repetition is avoided here.

The different dimensions that define ethnicity can be summarized as; kinship, religion, language, shared territory, nationality, and physical appearance (Bulmer 1996). In principle one could accurately classify a person into an ethnic group if these six dimensions were to be measured separately. This conclusion has been reached by several researchers in ethnic inequalities in health, that call investigators to use a range of variables instead of just one summary measure. Amongst common identity variables now available are: language, religion, country of birth, family origins, and length of residence (Bhopal 2004; Gerrish 2000; McAuley et al. 1996). Physical appearance seems to be a much more sensitive aspect to ask about, and even more to classify.

Even the trend in national censuses is now towards measuring these different dimensions separately. In the UK, in addition to the traditional country of birth question, an ethnicity question was introduced in 1991, a religion question in 2001, and questions on language spoken at home, passports held, national identity, and year of arrival were introduced in the 2011 Census (Mateos 2014). The recent collection of these new variables in the UK Census will not only provide a richer insight into ethnic minorities in Britain, for example allowing public services to be better targeted to different languages, but it will also allow for a key aspect of an individual's identity to be further revealed.

2.3 Conclusion: Ethnicity, Populations, Languages and Names

The evidence of ethnic and racial inequalities in most multicultural societies has grown strongly in the last decades (Finney and Simpson 2009; Mason 2003; Nazroo 2003). One of the aspects in which such inequalities are manifest is in its spatial dimension, with debates about ethnic residential segregation and the "ghettoisation of society" having acquired special prominence in the public debate of recent years (Dorling 2005; Finney and Simpson 2009; Phillips 2005). Although a range of diverse and intertwined factors for such ethnic inequalities has been identified, research has fallen short of unveiling the true interaction between such factors, especially at the local micro-level (Karlsen et al. 2002). The main problem has been a lack of availability of ethnicity data at sufficient quality and level of disaggregation, and an absence of adequate methods to interpret the problematic nature of measuring different ontologies of ethnicity (Mateos 2011; Mateos et al. 2009).

Therefore, new methods are required in the analysis of ethnic inequality in increasingly diverse populations and neighbourhoods, which are capable of being adapted to rapid changes in international migration and ethnic group formation processes. Such improved methods will prove key in informing policy to reduce ethnic inequalities, produce and maintain accurate population statistics and plan for the future complex needs of our societies and cities.

This book aims to contribute to such methodological need. It contends that there is a strong relationship between the ethnic identities of human groups and their mother languages or those of their ancestors, and that an indication of these can be revealed by the analysis of personal name origins. This is the cornerstone of the methodological innovation that this book aims to contribute: developing a new classification of populations and neighbourhoods along the multidimensional aspects of collective identity, through the cultural, ethnic and linguistic origins of personal names.

If this hypothesis can be proved correct and a suitable methodology can be developed for the purpose of studying ethnic group distribution at neighbourhood level, this research may be invaluable in overcoming the problems arising from ethnicity being measured as a single variable, the difficulties in classifying and generalising about ethnicity, the lack of data between censuses, and the coarse categorisations that census-type surveys adopt.

References

Afshari R, Bhopal R (2002) Changing pattern of use of 'ethnicity' and 'race' in scientific literature. Int J Epidemiol 31:1074–1076

Agyemang C, Bhopal R, Bruijnzeels M (2005) Negro, Black, Black African, African Caribbean, African American or what? Labelling African origin populations in the health arena in the 21st century. J Epidemiol Community Health 59(12):1014–1018

American Sociological Association (2002) Statement of the American Sociological Association on the importance of collecting data and doing social scientific research on race. Adopted by the elected Council of the American Sociological Association. Available at http://www.asanet.org/galleries/default-file/asa_race_statement.pdf. Accessed 20 Sept 2006

Arday SL, Arday DR, Monroe S, Zhang J (2000) HCFA's racial and ethnic data: current accuracy and recent improvements. Health Care Financ Rev 21(4):107–116

Aspinall PJ (2000) The new 2001 census question set on cultural characteristics: is it useful for the monitoring of the health status of people from ethnic groups in Britain? Ethn Health 5(1):33–40

Aspinall PJ (2002) Collective terminology to describe the minority ethnic population: the persistence of confusion and ambiguity in usage. Sociology 36(4):803–816

Aspinall PJ (2005) The operationalization of race and ethnicity concepts in medical classification systems: issues of validity and utility. Health Informatics J 11(4):259–274

Aspinall PJ (2007) Approaches to developing an improved cross-national understanding of concepts and terms relating to ethnicity and race. Int Sociol 22(1):41–70

Aspinall PJ (2009) The future of ethnicity classifications. J Ethnic Migrat Stud 35(9):1417–1435

Barrier NG (1981) The census in British India. Manohav, New Delhi

Bhopal R (1997) Is research into ethnicity and health racist, unsound, or important science? BMJ 314(7096):1751–1756

Bhopal R (2004) Glossary of terms relating to ethnicity and race: for reflection and debate. J Epidemiol Community Health 58(6):441–445

Bhopal R (2007) Ethnicity, race, and health in multicultural societies. Oxford University Press, Oxford

Bhopal R, Donaldson L (1998) White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. Am J Public Health 88(9):1303–1307

Brubaker R (2004) Ethnicity without groups. Harvard University Press, London

Bulmer M (1996) The ethnic group question in the 1991 census of population. In: Coleman D, Salt J (eds) Ethnicity in the 1991 census. Demographic characterisitics of the ethnic minority populations, vol 1. Office for National Statistics, HMSO, London, pp xi–xxix

Castells M (1997) The power of identity. Information age: economy, society and culture, vol 2. Blackwell, Oxford

Castles S, Miller MJ (2003) The age of migration, 3rd edn. Palgrave Macmillan, Basingstoke

Cavalli-Sforza LL (1997) Genes, peoples, and languages. Proc Natl Acad Sci USA 94(15): 7719–7724

Chapman RR, Berggren JR (2005) Radical contextualization: contributions to an anthropology of racial/ethnic health disparities. Health 9(2):145–167

Choi KH, Sakamoto A (2005) Who is Hispanic? Hispanic ethnic identity among African Americans, Asian Americans, and Whites. PRC Working Paper Series. Rep. No. 04-05-07. Population Research Centre, University of Texas at Austin. Available at http://www.prc.utexas.edu/working_papers/wp_pdf/04-05-07.pdf. Accessed 22 Feb 2005

Christopher AJ (2002) "To define the indefinable": population classification and the census in South Africa. Area 34(4):401–408

Coleman D, Salt J (eds) (1996) Ethnicity in the 1991 census. Demographic characteristics of the ethnic minority populations, vol 1. Office for National Statistics, HMSO, London

Comstock RD, Castillo EM, Lindsay SP (2004) Four-year review of the use of race and ethnicity in epidemiologic and public health research. Am J Epidemiol 159(6):611–619

Connolly H, Gardener D (2005) Who are the 'Other' ethnic groups? Social and welfare reports. Office for National Satistics, London. Available at http://www.statistics.gov.uk/articles/nojournal/other_ethnicgroups.pdf. Accessed 27 Jan 2006

Couzin J (2005) To what extent are genetic variation and personal health linked? Science 309:81

Dorling D (2005) Why Trevor is wrong about race ghettos. The Observer, Sunday, 25 September. Available at http://www.guardian.co.uk/race/story/0,11374,1577790,00.html

Eriksen TH (2002) Ethnicity and nationalism, 2nd edn. Pluto Press, London

Finney N, Simpson L (2009) 'Sleepwalking into segregation'? Challenging myths about race and migration. Policy Press, Bristol

Frazier JW, Margai FM, Tettey-Fio E (2003) Race and place: equity issues in urban America. Westview Press, Boulder, CO

Nature Genetics (2001) Editorial. Genes, drugs and race. Nat Genet 29:265–269

Gerrish K (2000) Researching ethnic diversity in the British NHS: methodological and practical concerns. J Adv Nurs 31:918–925

Gill P, Bhopal R, Wild S, Kai J (2005) Limitations and potential of country of birth as proxy for ethnic group. Br Med J 330(7484):196

Gould S (1984) The mismeasure of man. Pelican, London

Graves JLJ (2002) The emperor's new clothes. Biological theories of race at the millennium. Rutgers University Press, New Brunswick, NJ

Harding S, Dews H, Simpson S (1999) The potential to identify South Asians using a computerised algorithm to classify names. Popul Trends 97:46–50

Harvey D (2005) A brief history of neoliberalism. Oxford University Press, New York

Hitler A (1925) Mein Kampf. Secker and Warburg, Munich

HM Government (1976) Race Relations Act 1976. HMSO, London

Howard D, Hopkins PE (2005) Editorial: race, religion and the census. Popul Space Place 11(2):69–74

Kahn J (2005) Misreading race and genomics after BiDil. Nat Genet 37(7):655–656

Karlsen S, Nazroo JY, Stephenson R (2002) Ethnicity, environment and health: putting ethnic inequalities in health in their place. Soc Sci Med 55(9):1647–1661

Kertzer DI, Arel D (2002) Census and identity. The politics of race, ethnicity, and language in national censuses. Cambridge University Press, Cambridge

Marmot M, Adelstein A, Bulusu L (1984) Immigrant mortality in England and Wales 1970–78: causes of death by country of birth. OPCS. Her Majesty's Stationery Office, London

Mason D (2003) Explaining ethnic differences: changing patterns of disadvantage in Britain. Policy Press, Bristol

Mateos P (2011) Uncertain segregation: the challenge of defining and measuring ethnicity in segregation studies. Built Environ 37(2):226–238

Mateos P (2014) The international comparability of ethnicity classifications and its consequences for segregation Studies. In: Lloyd C, Shuttleworth I, Wong D (eds) Social-spatial segregation: concepts, processes and outcomes. Policy Press, Bristol, UK

Mateos P, Singleton A, Longley P (2009) Uncertainty in the analysis of ethnicity classifications: issues of extent and aggregation of ethnic groups. J Ethnic Migrat Stud 35(9):1437–1460

McAuley J, De Souza L, Sharma V, Robinson I, Main CJ et al (1996) Self defined ethnicity is unhelpful. Br Med J 313(7054):425b–426b

M'charek A (2005) The human genome diversity project. Cambridge University Press, Cambridge

Mitchell SA (1868) The new primary geography. E.H. Butler & Co., Philadelphia

Mitchell R, Shaw M, Dorling D (2000) Inequalities in life and death: what if Britain were more equal? Policy Press, Bristol

Modood T (2005) Multicultural politics: racism, ethnicity and muslims in Britain. Edimburg University Press, Edimburg

Morning A (2008) Ethnic classification in global perspective: a cross-national survey of the 2000 census round. Popul Res Policy Rev 27(2):239–272

Nazroo J (2003) Patterns of and explanations for ethnic inequalities in health. In: Mason D (ed) Explaining ethnic differences: changing patterns of disadvantage in Britain. Policy Press, Bristol

Nobles M (2000) Shades of citizenship: race and the census in modern politics. Stanford University Press, Stanford

Office for Management and Budget (1978) Directive No. 15: racial and ethnic standards for federal statistics and administrative reporting. *Federal Register* 43(May 4):19269

Office for National Statistics (2003) Ethnic group statistics: a guide for the collection and classification of data. Available at http://www.statistics.gov.uk/about/ethnic_group_statistics/downloads/ethnic_group_statistics.pdf. Accessed 13 Feb 2006

Olson S (2002) Mapping human history: genes, race, and our common origins. First Mariner Books, New York

Oppenheimer GM (2001) Paradigm lost: race, ethnicity, and the search for a new population taxonomy. Am J Public Health 91(7):1049–1055

Peach C (1996) Ethnicity in the 1991 census. The ethnic minorities of Great Britain, vol 2. Office for National Statistics, HMSO, London

Peach C (1999) Social geography. Progr Hum Geogr 23(2):282–288

Peach C (2000) Discovering white ethnicity and parachuting plurality. Progr Hum Geogr 24(4): 620–626

Pearce N, Foliaki S, Sporle A, Cunningham C (2004) Genetics, race, ethnicity, and health. Br Med J 328:1070–1072

Pfeffer N (1998) Theories in health care and research: theories of race, ethnicity and culture. Br Med J 317(7169):1381–1384

Phillips T (2005) After 7/7: sleepwalking to segregation. Speech given to Manchester Council for Community Relations, Manchester. Available at http://www.cre.gov.uk/Default.aspx.LocID-0hgnew07s.RefLocID-0hg00900c002.Lang-EN.htm. Accessed 4 Mar 2006

Rankin J, Bhopal R (1999) Current census categories are not a good match for identity. Br Med J 318(7199):1696

Renan E (1990 [1882]) What is a nation. In: Bhabha H (ed) Nation and narration. Routledge, London, pp 7–19

Science (2005) So much more to know…. Science 309(5731):78–102

Senior PA, Bhopal R (1994) Ethnicity as a variable in epidemiological research. Br Med J 309(6950):327–330

Skerry P (2000) Counting on the census? Race, group identity, and the evasion of politics. Brookings Institution Press, Washington

University of Cambridge (2004) English Dictionary. Cambridge University Press, Cambridge

Weber M (1980 [1921]) Wirtschaft und Gesellschaft (Eng. Tr. Economy and Society). Mohr, Tübingen

Whitehead M (1992) The health divide. In: Townsend P, Whitehead M, Davidson N (eds) Inequalities in health: the black report and the health divide. Penguin, London

Wild S, McKeigue P (1997) Cross sectional analysis of mortality by country of birth in England and Wales, 1970–92. Br Med J 314(7082):705–710

Names, Ethnicity and Populations
Tracing Identity in Space
Mateos, P.