 Approaches to Developing an Improved Cross-National Understanding of Concepts and Terms Relating to Ethnicity and Race

Peter J. Aspinall
University of Kent

abstract: Investigators from the fields of comparative social and epidemiological research have identified the need for an improved cross-national understanding of the concepts and terms relating to ethnicity and race. Suggestions have included the harmonization in surveys of variables like ethnicity and religion in a comparative European context and an internationally applicable and agreed glossary of terms relating to ethnicity and race. Pleas have been entered for work towards such goals, involving statistical offices and institutions in the European Union and bodies like the World Health Organization and International Epidemiological Association. This article examines how the conceptual bases of this terminology, issues of geographical specificity and the problem of which terms merit recognition impact on these goals. Different approaches to improving our understanding of this terminology in a cross-national context are explored. Given that the meanings of concepts and terms in the field of ethnicity and race invariably can only be understood in their national context of use – which is frequently layered, manifold, subtle and complex – an approach that explores the connotative reach of the different concepts and terms within this context is needed. Functional equivalence is more likely to be achieved by harmonization than the systematization of such knowledge through the economical form of a glossary of synthetic analytical terminology. However, given the socially and psychologically driven nature of ethnicity as a ‘global’ concept, harmonization may only be successful when limited to its multiple dimensions.

keywords: census • concept • ethnicity • race • terminology
Introduction

Investigators have identified the need for an improved cross-national understanding of the concepts and terms relating to ethnicity and race. Impetus has come from comparative social researchers working on cross-national comparisons of surveys and statistics within a European context. While much progress has been made in developing valid instruments with functional equivalence that allow for such comparative analysis (for example, for occupation and education), there are several single variables for which such standardized measures are not yet available. Hoffmeyer-Zlotnik (2003) and collaborators have suggested tools for the harmonization across European countries of some of these variables, including race and ethnicity. Lambert (2004) has also pursued the utility in the survey method of cross-nationally comparative analyses of ethnic differences and analytical strategies that might meet such need. A concern among epidemiologists about making meaningful comparisons in an international context of findings in the field of ethnicity and health has established a similar momentum. Bhopal (2004) has recently presented a set of terms relating to ethnicity and race derived mainly from the UK population context and invited debate on, and help in resolving, the problems of developing an internationally applicable glossary, which he first argued was needed in the early 1990s (Bhopal, 1992; Bhopal et al., 1991). This echoes a wider call from epidemiologists who have expressed similar concerns about the lexicon. Workers in other disciplines, including political scientists (Bird, 2003) and criminologists (Junger-Tas, 2001), have also reported difficulties in undertaking comparative cross-national research in the field of race and ethnicity.

The substantial body of work across research fields that explores the complex issues of the use of ethnicity and race and how they might be addressed (Aspinall, 2001; Modood et al., 1997;) lays a solid foundation for the development of an improved understanding of these concepts and terms in a cross-national context. The current difficulties in meeting this challenge include the range of meanings attributed to the terms ethnicity and race and recognition that many of the specific terms for individual groups need to be defined with respect to the different conceptual bases, notably, those of self- or group identification and social categorization. Many concepts and specific labels have national, regional or other geographical specificity, making any read-across of the terminology used for ethnic/racial groups fraught with difficulties. Unpacking the concept of racial or ethnic group into its several dimensions raises theoretical as well as methodological issues, including the validity of the resultant constructs as group measures.
In addressing the use of ethnicity in cross-national, comparative social research, this article explores the conceptual bases of the terminology – notably, ‘ethnicity’ and ‘race’ and ‘social categories’ and ‘group identities’ – and differences in internal and external definition with respect to the examples of ‘Asian’ and ‘White’ in the US and Britain. The consequences of categorization are examined, including conflictual categories and the denial of recognition within the context of official classification. Having described some of the difficulties in the conceptualization and measurement of ethnicity within national contexts, the article proceeds to investigate approaches to harmonizing ethnic/racial terminology cross-nationally. These include semantic locality, input and output harmonization, and the novel approach of the numerical indexing of different ethnicity categories within each country developed by Lambert (2004). The focus of the discussion of operationalizing ethnicity is primarily that of the construction of cross-national standardized instruments or measures in the survey setting rather than on qualitative research (such as phenomenology, symbolic interaction and ethnomethodology). More specifically, the article is concerned with instruments for comparative research with ethnicity as an independent variable.

However, as International Sociology’s recent special issue on comparative research has shown (Vol. 21, No. 5, 2006), a range of methodologies can be exploited that combine features of the case-oriented and the variable-oriented approach and provide a bridge between qualitative and quantitative analyses. In particular, the call by Ragin (2006) to focus greater effort on how populations are ‘constituted’, with respect to their degree of theoretical content, may, of itself, lead to the better conceptualization of independent variables like ethnicity and social class. For example, in much quantitative social research, social class is treated as a confounder of the relationship between ethnicity and various outcomes, something to be mechanistically ‘controlled for’. In reality, however, both are interrelated and mutually implicated: an investigation of how such categories are constructed, operationalized and feed back into and modify each other’s impact is likely to enrich the analysis of causal complexity.

The Conceptual Bases of the Terminology

‘Ethnicity’ and ‘Race’

One of the enduring methodological problems in comparative research is that of construct equivalence, that is, ‘the instance where the instrument measures the same latent trait across all groups, or nations, or cultures’ (Mills et al., 2006). In the case of ‘ethnicity’ and ‘race’ there are formidable barriers to building such cross-national comparative categorizations as a wide range of concepts are now used in describing the ethnic/racial composition.
of populations, such as ethnic group, ethnic origin, sociocultural group, race, racial group, ancestry, family origins, national origins and others. These concepts and the related specific labels for different groups have evolved within the context of distinct processes of ethnogenesis specific to individual countries, including colonialism and its legacy, migration, racism and discrimination. Consequently, a key challenge to the development of a cross-nationally agreed understanding will be that of mapping the meanings of these concepts and terms which are frequently nuanced and inconsistently used. Moreover, while officialdom (governments and census or statistical agencies) offer definitions for this lexicon, they may differ from those used by members of the academy, and by the people thus described.

The term ‘race’ illustrates these tensions. While ‘ethnicity’ is increasingly replacing ‘race’ as the preferred term in much official, public and lay discourse, the latter remains in wide use. In the US, the Census Bureau termed its census question ‘race’ in the 1990 and 2000 censuses, describing the categories as a mix of both racial and national origin or sociocultural groups, and ‘race’ is the salient term in the wider population. In Canada, however, the statistics agency has eschewed ‘race’ for the terminology of ‘population group’ and ‘visible minorities’, this translation of colour into the language of visibility being viewed by some as a way of avoiding issues of racism (Stasiulis, 1991). The term widely used by officialdom in Britain is that of ‘ethnic group’, although such groups are clearly racialized, and ‘race’ is a focus in some branches of government and in official ‘race relations’. In the international scientific literature ‘ethnicity’ is on the ascendant although, again, the specificities of national contexts and traditions drive the debate in the broader sociological arena. Representative of a body of opinion on the use of the term ‘race’ in the US, for example, is the view that: ‘Given the burden of US racism, the racialized ethnicity of, for instance, black Americans cannot be erased by the popular force of mere semantic ethnicization, the upgrading of black social identity by the “African American” self definition’ (Harrison, 1994).

Moreover, the understanding of these various concepts and terms by the wider society, including the communities they describe, affect their utility as descriptors. The literature is limited on the extent to which people perceive a difference between concepts like ‘race’, ‘ethnic group’ and ‘national origins’ or view them as part of the same semantic domain. Clearly, such perceptions might affect the quality of data collected in demographic and social surveys, one of the key sources for investigating ethnic/racial disparities. Cognitive research undertaken by the US Census Bureau for the Current Population Survey Supplement (McKay and de la Puente, 1995) revealed that respondents frequently were unable to separate race, ethnicity and ancestry. In Britain, some participants in cognitive and focus group research saw a difference between ‘ethnic origin’ and ‘ethnic group’ while
others did not (Rainford, 1997; Mortimer and White, 1996). However, ‘ancestry’ was a term that many found confusing and imprecise. A further body of findings also suggests that, in the wider society, concepts like ‘race’ or ‘racial group’ are interpreted in a variety of ways, including biological difference, a label of social value and a synonym for culture.

Social Categories and Group Identities

In addition to debates about the utility of the concepts of race and ethnicity, there are also important differences with respect to how collective identities are theorized. Three distinct types of collectivity have been distinguished as part of this conceptual framework. First, there is official categorization, that is, a collective external definition or social categorization. A major part of the vocabulary for ethnicity and race, as used by government and in wider policy settings, comprises such social categories. An important but not the only source of these categories is the decennial national censuses, whose classifications frequently become the national gold standard. They are usually established by census and national statistical agencies upon the basis of phenomenal distinctions judged appropriate by a community of scientific and other observers, their adequacy being ‘ultimately a matter of the extent to which they contribute to the construction of cross-culturally testable hypotheses and theories’ (Harris et al., 1993). Second, collective external definition by the wider society is another form of such categorization. However, there is a substantial body of evidence that indicates that such external recognition is not a satisfactory means for distinguishing ethnic groups, except perhaps at the broadest, pan-ethnic level, and it has been little used in recent decades. For example, the US Statistical Policy Directive No. 15, the standard for ethnic/racial group classification, adopted in 1977, indicated that, for persons who identified with more than one race, the single racial category that most closely reflects the individual’s recognition in his or her community should be used. Such societal assignment now tends to be eschewed by officialdom. There remain only a few contexts in which it can be justified, including the specific need to obtain some measure of socially perceived ethnicity or the way ethnicity is construed in everyday social interaction.

A third type of collective social identity is that of self-identification with respect to a recognizable collective identity. This collective internal definition or group identification is established by eliciting respondents’ categorizations of themselves or of their family members through their own self-descriptions or terms that they regard as appropriate, whether or not observers find them to be ambiguous or contradictory. In practice, these two processes – social categorization by scientific observers and the wider society and internal definition or group identification – are mutually entailed and
feed back upon each other. However, while social categories frequently validate a group’s internal definition, this is not always the case and there are examples where such processes have given rise to different and conflictual vocabularies. Scientific investigators have an interest in both social categorization and group identification (Aspinall, 1997; Nazroo, 1998), although unprompted self- or group identification has been relatively neglected. While its utility has been challenged on the basis of the resulting selective attribution – and consequent lack of stability – in self-assigned descriptions (McAuley et al., 1996; Reijneveld, 1998) and problems of category derivation from the myriad of responses, there has been a marked shift to this ‘cognitive tradition’ in recent years. Smith (2002) considers the distinction an important one, preferring to reserve the term ‘ethnic group’ for self-ascriptions by group members and ‘status group’ (after Weber) for the social identification of a group by such criteria as family origins.

In some European countries there has been a stronger tradition of interpreting ethnicity both in terms of its measure as a self-identity and through multidimensional measurement, including country of birth, parents’ country of birth, migration status, language(s) spoken, religion, citizenship and nationality. The importance of nationally specific contexts is illustrated by France and Belgium, for example, where the state has declined to demarcate its citizens by either race or ethnicity. The complexity and multiplicity of national origin groups across Europe makes any attempt to draw cross-national comparisons fraught with difficulty. Kymlicka (1995) has attempted to distinguish between ‘national minorities’ – those long-standing minority groups whose presence in a state might have resulted from colonialism and its legacy and migration, and polyethnic minorities, groups characterized by more recent arrival including asylum seekers and refugees. The latter can frequently be identified by country of ethnic or family origins, the mapping of such groups in a cross-national context being facilitated by use of a geographical referent. However, when data are collected on the basis of self-description, the plethora and, in some cases, uniqueness of the terms used makes the identification of minority ethnic groups much more problematic. They may include, for example, the use of bicultural terms like ‘Scottish-Muslim’ where the term gives recognition to the influence of the host culture. Caglar (1997) cites other examples of these hyphenated identities – German-Turks, British-Pakistanis, French-Algerians and European-Muslims – in which hyphenation generally privileges nationality or territorialized religion. These difficulties are compounded in the case of multiple allegiances and attachments, as is frequently found in those of mixed race or heritage. Such subjects may choose to identify in an extremely diverse range of ways, frequently encompassing multiple groups and, in some cases, also nationalities (Mahtani, 2001; Rockquemore, 1998).
Internal and External Definition: 
Examples of the Process of Naming 
in Britain and the US

The way in which differences in history, migration processes, social structure, culture and racialization across the Atlantic divide have contributed to different traditions in the conceptualization of ethnic/racial groups – and, indeed, of research practice – illustrates the problems in developing cross-national vocabularies. Several labels for these different social categories are identical – such as the terms ‘Asian’ and ‘White’ – and it is clear that the meanings of these categories can only be elicited by recourse to definitions in context.

The Example of ‘Asian’

The term ‘Asian’ is a social category used by officialdom, the research community and in the wider public discourse – as well as being a group identifier (a self descriptor for those thus described) – in Britain, the US and Canada. In Britain, the term ‘Asian’ may refer to those who have origins in the Indian subcontinent or in continental Asia (Aspinall, 2003a). Even the 2001 Census treats the term differently across the home countries, its usage in England and Wales being that of the Indian subcontinent but, in Scotland, inclusive of Chinese. Such usage mirrors how the term is used in official reports and the research literature.

There are, too, differences with respect to popular conceptualization: how the wider public interpret such categories in data collection affects their utility and validity. For example, the Office for National Statistics clearly intended ‘Asian’ in the England and Wales 2001 Census to refer to Indian subcontinent, as revealed by the composition of codes used for open response options. However, an examination of countries of birth of those who ticked the 2001 Census ‘White and Asian’ mixed category suggests that many interpreted it otherwise: while 79.7 percent of the 189,016 persons were born in Europe (mainly the UK), 4.3 percent were born in the Middle East, 5.1 percent in the Far East and 6.8 percent in South Asia (Office for National Statistics, 2003). Among the ‘Other Asian’ group listed under the ‘Asian or Asian British’ option, the respective percentages for these categories were 15.9 percent, 2.0 percent and 36.6 percent (respondents in this case could have chosen one of the two options under the ‘Chinese or other ethnic group’). Research studies also point to such differences in interpretation based on the population density of the group. ‘Asian’ appears to have some saliency as a self descriptor for those of Indian subcontinent origins (reflecting the wider popular conceptualization) in areas where South Asian groups comprise a significant proportion of the population, as in Leicester (Pringle and Rothera, 1995), but none in areas where the group is
small in size (as in Tyneside) (Rankin and Bhopal, 1999). Moreover, focus group studies among those with origins in the Indian subcontinent have revealed that ‘Asian’ is accepted purely as a description, including one they would use of themselves, but not a positive identity. However, ‘South Asian’ (a popular term in official and research contexts) is a term that such participants do not understand and do not apply to the Indian subcontinent

Table 1  Official Terms for, and Definitions of, Some ‘Pan-Ethnic’ and Specific Groups in Britain, USA and Canada

<table>
<thead>
<tr>
<th>Group</th>
<th>Britain</th>
<th>United States</th>
<th>Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘White’</td>
<td>In the national (the three home country) 2001 censuses ‘White’ category, as persons with cultural background or origins in European countries.</td>
<td>In the 2000 Census ‘White’ category – as a person having origins in any of the original peoples of Europe, the Middle East or North Africa.</td>
<td>In the 2001 Census ‘White’ category, as persons with origins in European countries (separate categories were provided for ‘West Asian’ and ‘Arab’).</td>
</tr>
<tr>
<td>‘Asian’</td>
<td>In the England and Wales 2001 Census category ‘Asian or Asian British’ – as persons of Indian subcontinent cultural background only; in the Scotland 2001 Census category ‘Asian, Asian Scottish or Asian British’ – as continental Asian, as referenced by Indian subcontinent, Chinese and ‘Any other Asian background’ options.</td>
<td>In the 2000 Census seven Asian groups, including ‘Asian Indian’ and ‘Other Asian – print race’ – as a person having origins in any of the original peoples of the Far East, Southeast Asia or the Indian subcontinent.</td>
<td>In the 2001 Census ‘Population Group’ question, as the categories ‘South Asian’, ‘Southeast Asian’ and ‘West Asian’.</td>
</tr>
<tr>
<td>(South Asian) ‘Indian’</td>
<td>‘Indian’</td>
<td>‘Asian Indian’</td>
<td>‘South Asian’ only in Population Group question; suggested terms in open response ethnic origin/ancestry question: ‘East Indian from India’, ‘East Indian from Guyana’ or the specific group, such as ‘Punjabi’ or ‘Tamil’.</td>
</tr>
</tbody>
</table>
This clearly polysemic nature of the term ‘Asian’ – as having a range of meanings that vary with the contexts of use – thus creates problems of interpretation even in a single country.

Difficulties with the use of this vocabulary are also found in North America (Table 1). The terms ‘Asian’ and ‘Asian American(s)’ are widely used as social categories in the US. In the 1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, ‘Asian’ is defined as ‘a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam’ (OMB, 1997). The 2000 US Census similarly uses the term ‘Asian’ in the set ‘Asian Indian’, ‘Chinese’, ‘Filipino’, ‘Japanese’, ‘Korean’, ‘Vietnamese’ and a residual ‘Other Asian’ free text field. This is also the use of the term ‘Asian American’ in the US National Library of Medicine’s 2004 MeSH subtree for ‘Population Groups’ (Schulman, 2003). However, as in Britain, there are differences in popular conceptualization, such collective terms being infrequently used as self-descriptors. Survey evidence shows that ‘Asian American’ has only limited saliency among the Chinese, Korean, Vietnamese, Japanese, Filipino and South Asian origin groups – ranging from 12 percent (Chinese) to 23 percent (South Asian) – and ‘Asian’ virtually none (range 2–7 percent) (Lien et al., 2003). Terms like ‘Chinese American’, ‘Korean American’, etc., or simply one’s ethnic origin (e.g. Chinese, Vietnamese, etc.) predominate, although in this survey most Japanese (41 percent) simply identified as ‘American’, a term not popular with the other groups.

The experience of Canada shows that, within the North American continent, national boundaries are also terminological frontiers. The 2001 Canadian Census question on ‘Population Group’ (as termed in the Census Guide) uses ‘Asian’ in the categories ‘South Asian’, ‘Southeast Asian’ and ‘West Asian’, the last term having no saliency in Britain or the US; these three options were the only ones among the 11 predesignated categories where examples were given in the question, perhaps indicating the need to define terms that carry a degree of ambiguity. Thus, at the level of specific groups, the choice of options varies across the continent. In the US, a person with origins in India is offered the ‘Asian Indian’ option (this option relating specifically to India as the question instruction and guide in the 1990 Census directed ‘Pakistanis’ and ‘Bangladeshis’ to the free text ‘Other Asian’ field), but in Canada the ‘South Asian’ option. In the free text census questions on ethnic origin/ancestry respondents had a further opportunity to self-assign, the example of ‘East Indian’ being offered in the Canadian question, although the guide provided the instruction to respondents: ‘For persons of East Indian or South Asian origins, report a specific group. Do not report “Indian”. For example, report “East Indian from India”, “East Indian from Guyana”, or...
indicate the specific group, such as “Punjabi” or “Tamil” (Statistics Canada, 2002). Again, ‘East Indian’ is a term that has no saliency in the US or Britain.

**The Example of ‘White’**

‘White’ is another widely used social category that is geographically specific in its official meaning and also reveals differences in the traditions of popular conceptualization. ‘White’ in Britain is widely understood to refer to people with European ancestral origins and is so defined by the Office for National Statistics in terms of its full coding list for the ‘White’ ethnic group in the 2001 Census question. However, it is not necessarily so interpreted by respondents in such surveys. For example, it is notable that, in the 2001 England and Wales Census, of the 217,352 respondents born in the Middle East and 68,715 born in North Africa, 60.1 percent and 50.9 percent, respectively, ticked one of the ‘White’ category cultural background options (Office for National Statistics, 2003). It is likely, therefore, that in data collections the ‘White’ category is substantially more heterogeneous than official agencies intend, justifying concerns that it should be the subject of greater focus (Aspinall, 1998; Bhopal and Donaldson, 1998). In the US, by contrast, the 1997 revisions to the statistical standard define ‘White’ as ‘a person having origins in any of the original peoples of Europe, the Middle East, or North Africa’. Synonyms for the term ‘White’ are equally problematic. The category ‘Caucasian’ has limited saliency as a group identifier in the US, chosen as the preferred term by just 16.5 percent of Whites in a population-based survey (Tucker et al., 1996), but by far fewer in Britain in a similar survey (just 3.5 percent of persons who described themselves as ‘White’ when prompted with the 1991 Census question) (Pringle and Rothera, 1995). Yet this term is the most frequently used and proportionately increasing in medical literature databases, even those with a European as well as North American emphasis (Aspinall, 1998). In the US, there is even some regionalization in the use of ‘White’ and its synonyms, ‘Anglo’ being used by only 1 percent of the population nationally (Tucker et al., 1996), but by a much higher proportion in the southwest of the United States.

The foregoing discussion has illustrated the difficulties that arise in endeavouring to harmonize terminology for broad-based racial or ‘pan-ethnic’ groupings. Nations are shaped by different long-term historical processes that include their colonial past, culture and social and political context. These empirical differences, in turn, give rise to different traditions in the popular conceptualization of race and ethnicity, ways in which officialdom interprets these and the ‘constructs’ it offers, and in research practice. Such difficulties are magnified when cross-national comparisons are required not just for the major national ethnic groupings like ‘Asian’ or ‘White’ but the full range of categories that are common in
racial/ethnic classifications, now frequently numbering 15 or more. This proliferation in category options has resulted from the growing ethnic and racial diversity of populations within different nations, arising from increasing rates of interethnic unions, developing global patterns of migration and diasporas, and changes in the fashion of particular labels and group identifiers. Moreover, how successful census and statistical agencies are in tracking such diversity is limited by their knowledge of the emergence of new ethnic groups (or ways in which they choose to be identified) and, in some cases, by the politicization of ethnic identity within nation-states.

The Consequences of Categorization

The categorization of ethnic groups, especially by officialdom, is not just a matter of social classification practices: it has consequences. Moreover, such consequences extend to both individuals and populations. In the bureaucracy of the modern state, ethnic group data on individuals is frequently collected in a wide range of settings, including education, the labour market and health and social care. Around every 10 years the national population census asks for a set of personal information that often includes race or ethnic group, such information becoming the source of a wide range of official national statistics at the population level. These practices of social classification are used by the state for the distribution of resources and penalties in society and so are consequential for the lives of individuals and welfare of groups.

At an individual level, social categorization or definition by others can have a ‘labelling effect’ – whereby categories are mapped onto ethnic/racial stereotypes – which may result in one of two consequences. First, when the labelling is by others who have authority or power and occurs in an institutional social setting, the person may acquire or internalize the applied identity (Jenkins, 1996). Labelling may also produce resistance, even when only the capacity exists (as when, for example, people fail to complete an equal opportunities monitoring or census question on ethnicity on grounds of acceptability). What effect labelling has depends substantially on the extent to which such labels validate internal self-definitions and are judged to be positive or negative by the recipient. However, in terms of both the distribution of resources and exertion of penalties, its effects have been extensively documented in a number of areas that affect people’s social experiences and life chances, including education assessment, recruitment in the labour market, access to welfare benefits, allocation of social housing and provision of health care.

It is frequently at the population level that the practical consequences of categorization most clearly impact on ethnic groups. Here, it is officialdom,
particularly in the form of national census agencies, that perhaps has had the greatest capacity to generate such consequences. The decennial census has traditionally supplied much of the official vocabulary for different ethnic/racial groups and carries commensurate status and authority. It is recognized that census classifications are usually the outcome of complex processes of cognitive research, question testing and large-scale trials and so are readily adopted by the state as the standard. The census thus provides the ‘denominator’ for many measures, in the form of rates and ratios, and is frequently used in surveys as a basis for sampling, to adjust for survey non-response, and to provide weights for national-level estimates.

**Conflicting Categories**

In order to conduct a full enumeration, the census has to derive categories that include everyone in what is often an increasingly diverse population. Once again, the process of developing ethnic/racial classifications involves exploration of internal group definitions and the social categories of the state. How – and even if – groups get named in this process is consequential for their recognition, status and influence. While census development programmes increasingly involve community groups and endeavour to solicit and respond to their views, the adopted classifications are invariably contested on the grounds that the categories of officialdom conflict with group preferences. This may happen for a number of reasons. Group preferences frequently change in response to shifts in the views of community leaders and members, changes in fashion for particular terms and political pressures. Given the relative infrequency of censuses (decennial in the USA and Britain, quinquennial in Canada, for example), there may inevitably be a lag in the adoption of these preferences by officialdom. In other cases, the ‘process’ may be at fault. In Britain, for example, the development programmes are primarily reactive, involving the development of questions and their testing in both small and large trials, to the exclusion of the identification of salient terms in large, representative population samples (as happens in the US, for example; Tucker et al., 1996).

While broad official categories (such as ‘White’) can be mapped cross-nationally using official definitions (although these may not be consistent with popular conceptualizations), the role of the state can be idiosyncratic and, in some cases, contentious. Wider ethical issues arise when terms are used that the communities thus described do not find acceptable or result in an assignment of ethnicity that they regard as inaccurate or incomplete in meeting their needs, an outcome that may impact on such matters as data quality in the survey context and, consequently, administrative allocation. Cornell and Murphy (2002) point to the implicit dangers of the politics of
recognition: ‘In so far as it is articulated as a demand placed upon a dominant group and integrally tied to the substantiation of pre-given or fixed identity, it can easily mask or even reiterate cultural hierarchization associated with Euro-centrism.’ An example is the imposition of category options in the Brazilian census that overrode the cognitive system by which people expressed their ‘race–colour’ identities, illustrating the limitations of utilizing classification procedures deemed appropriate in the US national context (Harris et al., 1993). Equally consequential is the insistence by Cornell and Murphy that it should be people, as individuals and members of groups, who shape the identifications recognized by the state and not the state itself.

The state, of course, is usually a fully engaged partner as a constituter of social categories, rather than being either an impartial player or a demonic influence as in the past in Nazi Germany, South Africa in the apartheid era, or elsewhere. While ethnic/racial communities or groups are influencing the content of census questions to a much greater extent than in the past, especially in the US, the argument of Cornell and Murphy clearly supports a policy of more robust research on, and recognition of, group identifications in the different populations. When there is a failure in this approach, that is, when the state’s categories are not locally grounded, its role in the process of social categorization has been construed as instrumental. Commentators have described this outcome as that of ‘creating new kinds of persons for individuals to be’ (Brubaker et al., 2004), ‘nominating into existence’ (Goldberg, 1997), ‘making up people’ (Hacking, 1986) and producing ‘fictive unities’ (Werbner, 1990). Examples in the US, where the state as a categorizer is implicated, are the pan-ethnic terms ‘Asian or Pacific Islander (API)’ – used on the 1990 US Census form (and rapidly adopted as a ‘pan-ethnic’ group) but dropped in that for 2000 – and ‘Hispanic’ (used in the ‘Spanish/Hispanic origin’ question in 1990 and the ‘Spanish/Hispanic/Latino’ question in 2000). Such terms rapidly entered the lexicon of social and epidemiological research. ‘Hispanic’ had become the most popular collective term within the community by the mid-1990s, illustrating the suggestion by Peterson (1987) that ‘few things facilitate a category’s coalescence into a group so readily as its designation by an official body’. Again, there may be empirical differences between acceptance and popular conceptualization, it being more probably the case that, although only partially and situationally meaningful to the people thus labelled, such pan-ethnic terms are nonetheless employed by them in their dealings with officialdom.

How groups lose out or benefit in this category shuffle may be subtle. The use of the terms ‘Hispanic’, ‘Spanish’ and ‘Latino’ has been mentioned: some have opposed ‘Hispanic’ as an imposed terminology, while the popularity of ‘Latino’ is regional; yet others prefer to identify using national origin, such as ‘Cuban’ or ‘Mexican’. Differences in the count of
the population to which these term(s) are meant to apply and the number actually choosing the term(s) may be consequential for the overall count and, hence, resource distribution. The same argument applies with respect to ‘African American’ vs ‘Black’, research showing that the foreign-born black population from the Caribbean and Africa do not believe that ‘African American’ refers to them. Similar debates have focused on ‘Hawaiian’ vs ‘Native Hawaiian’ (the latter arguing that they were the original inhabitants of territories that the US acquired) and terms for those with origins in Guam (‘Chamorro’ competing with ‘Guamanian’). In Britain, young people of ‘Black Caribbean’ parentage born and educated in the country do not feel that the term references them, preferring to identify as ‘Black British’. When broad labels are used that encompass socioeconomically heterogeneous groups, those that are more disadvantaged frequently argue that their legitimate claims for resources are lost in aggregate data: indeed, it was not until 2001 that the ‘Irish’ group was separately distinguished within the ‘White’ category in the UK censuses.

**Official Categorization and the Denial of Recognition**

Given the pervasiveness in the modern state of administrative allocation based on population categories, the denial of official recognition perhaps demonstrates the power of naming at its most consequential. Many ethnic/racial groups are not so recognized, either because they are too small, difficult to enumerate or outside the state’s political imperatives. Arguably, cross-national collaboration on ethnic/racial concepts and terms may provide a more robust scrutiny of the population categories that the state chooses to recognize – especially with respect to national minorities, refugees and asylum seekers, economic migrants and other ethnic diasporas and transnational communities – by removing such issues from the context of the ideologies and politics of the nation-state and relocating them in the social realities of the postmodern and globalizing world. Examples are how ‘Arabs’, Gypsy Travellers and Roma and persons of mixed race or heritage are treated in national censuses and data collections.

The category of ‘Arab’/‘Middle Easterner’ is frequently omitted in censuses for methodological and political reasons. In the comprehensive review of US federal standards for racial/ethnic data, public comment supported having a separate category to provide the necessary data to monitor discrimination against this population, but such a category was not added. However, the government indicated that further research was needed to determine the best way to improve data on this population group. The US 2000 Census question on race also omitted the group, yet the written answers to the ancestry question on the long form showed that 1.2 million reported an Arab ancestry, up from 860,000 in 1990. This increase in the Arab population of nearly 40 percent during the 1990s
compared with a rise in the total US population of 13 percent (de la Cruz and Brittingham, 2003).

The ‘Arab’ group was also omitted from the UK 2001 censuses, including the detailed coding frame for written responses, the Office for National Statistics (2003) admitting that responses like ‘Arab’, ‘Iranian’, ‘Moroccan’, ‘Muslim’ and ‘Other Middle Eastern, e.g. Iraqi, Lebanese, Yemeni’ and ‘Other North African’ could not be systematically allocated to 2001 Census ethnic groups. Its candidature has been limited to the 1975–86 census field trials (including the 1983 ‘model’ question of the House of Commons Home Affairs Sub-Committee on Race Relations and Immigration) and recent research on population groups with origins in North Africa, indicating that terms like ‘Afro-Arab’ and ‘Arab-African’ may be acceptable (Elam et al., 2000). In Canada, however, where the concepts of diversity and the mosaic society and the wider discourse of multiculturalism have given rise to a more inclusive capture of groups, the 2001 Census question on population group included an ‘Arab’ category (split off from the ‘Arab/West Asian’ category in the 1996 Census). Office codes for written responses also include ‘North African Arab’ used in English-speaking Canada and ‘Maghrébin’, a group into which the North African Arab community has coalesced in Quebec. In France, by contrast, where the focus is upon nationality, officialdom categorizes the French-born children or grandchildren of Arab-speaking migrants as simply ‘French’ without further qualification (Bird, 2003).

Censuses and terminology guides also frequently omit Gypsy Travellers and Roma, a situation that, with respect to Britain, may have been less likely to have arisen in a context of cross-national European collaboration where the presence of such groups in the Balkans, Iberian countries and elsewhere is officially recognized. In the three UK 2001 censuses, only the Northern Ireland Census enumerated the group using the category ‘Irish Traveller’ (the 2002 Ireland Census also asked about membership of the ‘Irish Traveller Community’). These communities have now been given recognition – through the use of ‘Traveller of Irish heritage’ and ‘Gypsy/Roma’ categories – in the government’s pupil-level annual school censuses in England and Wales and have been included in tests of questions in Scotland for the 2011 Census. The substantial invisibility of these communities in official data (including lack of research on preferred group identifiers) and long neglect in public policy is of concern, given that their health status and social exclusion is the worst of any ethnic group.

The issue of ‘mixed race’ is only beginning to be addressed in official data and the wider sociological literature (Aspinall, 2003b; Ifekwunigwe, 2004), although methods of tabulating the population that were for the first time allowed to select more than one race in the US 2000 Census, including ‘fractional assignment’, have been the focus of much analytical
attention (Allen and Turner, 2001). However, as US government representatives admit: ‘important questions remain. How meaningful and predictive are the multiple race categories for health risk, health status and public health trends of these populations? Is the quality of the data derived from the multiple race tabulations equal to that for single race groups?’ (Lumpkin, 2001).

Such questions equally apply to the British context where a ‘Mixed’ category was introduced for the first time in the 2001 Census, that for England and Wales being broken down into four dual-heritage cultural background categories (thereby assuming mixed parentage at the expense of more complex and possibly multi-ethnic patterns of diversity). Operationally, these new and derivative categories also raise concerns about the identification of suitable sampling frames and the capture and representation of such populations in social survey samples.

Again, the gap between officialdom’s recognition and group identification has been problematic. In Britain, the census agency took the view in the 1980s that people of mixed descent often preferred not to be distinguished as a separate group (Sillitoe and White, 1992) and consequently made no provision, yet in the 1991 Census around 230,000 people eschewed the predesignated categories and wrote in a ‘mixed’ description (over 640,000 identifying as ‘mixed’ when specific categories were offered in the 2001 Census). Moreover, it is known that some of those who are, by parentage, black and white, for example, will opt in self-identification questions for ‘Black’, ‘White’ or some other group rather than a mixed description (Rockquemore, 1998). Indeed, most recent research shows that the way young mixed-race persons identify is extremely fluid and can differ by the setting in which the question is asked and by self-completion vs interview (Harris and Sim, 2002). When the ‘mark one or more’ instruction is used, the resulting number of permutations can be huge (even though around 93 percent of persons who reported two or more races in the US 2000 Census marked only two races; see Jones and Smith, 2001) without any obvious means of prioritizing the options selected or collapsing them to a more manageable set. Even the generic terminology for the ‘mixed’ category is contested. Some government departments in Britain favour ‘mixed heritage’, although ‘mixed race’ was the salient term among school pupils and parents (Tikly et al., 2004). In the US, ‘multiracial’ and ‘mixed race’ are the favoured group identifiers (Tucker et al., 1996), although government uses the term ‘two or more races population’ to avoid what it considers ambiguous labels.

The preceding analysis has revealed that the names given to ethnic/racial groups by officialdom – the so-called ‘nominal identities’ (Jenkins, 1996) – are often only meaningful within the context of individual nations. Population categories such as ‘Asian’ or ‘White’ refer to different populations in Britain and the US. Moreover, when these terms are used by groups as group identifiers or self-ascriptions, they frequently have yet different
meanings again and reference different constituencies at the population level. Further, it has shown that the labels used by officialdom are somewhat arbitrary and change over time. The needs of the modern state for statistical data to meet the bureaucratic practices of government, including resource allocation, dictate the construction of population categories into which people can conveniently be placed. In this process, the coherence and legitimacy of the categories are judged from the perspective of their utility for governance. This frequently involves standardization and parsimony at the expense of the diversity and diffuseness of the vocabulary of internal group definition. By such means identification becomes consequential for everyday individual and group experience. Terms are used that some groups and their members do not recognize or find acceptable, attracting misreporting or under-recording. In a system where resources are distributed on the basis of numerical representation, such groups incur a penalty. When the system of state categorization omits groups altogether, they are also excluded from processes of administrative allocation.

This specificity of terminology poses problems for the cross-national mapping of different ethnic and racial groups. Not only is there a proliferation of terms but these have meanings in local contexts. An examination of half a dozen academic journals found almost 20 terms in use to describe populations with origins in South Asia: ‘Asian’, ‘Asian descent’, ‘British Asian’, ‘South Asian’, ‘Indian subcontinent’, ‘Indo-Pakistani’, ‘Indian’, ‘British Indian’, ‘Indian Asian’, ‘Indo-Asian’, ‘Indian Punjabi’, ‘Pakistani’, ‘Pakistani Punjabi’, ‘British Pakistani’, ‘Bangladeshi’, ‘British Bangladeshi’, ‘Bengali’ and ‘British Bengali’. Such an assemblage of collective categories could be expanded if group identifiers were added. A more limited set of terms is proffered by officialdom as categories for an ethnic group and most have integrity (to some significant degree) as group identifiers. Some systematic measure needs to be taken of this diversity in terminology to identify which terms are most frequently used and thus merit critical evaluation and to highlight those that are conceptually or technically flawed, ambiguous, imprecise or racist.

Census concepts and terms are likely to continue to be a key source of ethnic/racial categories in social surveys and statistics and epidemiological research. However, since such surveys are frequently self-completion, in which respondents assign their ethnic or racial group, such categories need to be continually evaluated in terms of whether they are exhaustive (capturing all the main groups), whether the terms used are acceptable to those they describe, and whether they capture the populations they are intended to. Their level of success in these respects will determine the extent to which categories are representative of particular population groups (in terms of consistent capture) and, ultimately, their utility in the context of translating findings to a comparative cross-national context.
Much writing on race and ethnicity has been grounded in the historical specificity of particular national contexts. In recent years, however, increasing cooperation within political groupings like the EU has created a demand for comparative social research, including the use in cross-national contexts of valid demographic and socioeconomic instruments with functional equivalence. Similar demands have come from cross-national multi-centre research collaborations in the field of epidemiology, where a common vocabulary is needed. These demands of national statistical offices, agencies and others call for the exploration of ways of harmonizing single variables, including race and ethnicity, that have not been standardized. The dilemma for social researchers is that different nations differ empirically in terms of how their history, political trends, social structure, migration histories and culture have shaped the evolution of concepts in this field. These wider processes have, in turn, produced local situations with respect to official terminology for race/ethnicity, popular conceptualizations and research practice. The challenge thus becomes one of devising methods of collecting or reporting data that genuinely contribute to cross-national comparisons but also do not impugn these distinctive, nationally specific empirical conditions.

Many sociologists would argue that the particularity of racial/ethnic identity is such that it is not amenable to such processes of harmonization. Indeed, the view that it is the context-dependent and uncontrolled nature of discourse that allows language to work and evolve and new concepts to emerge, is probably a majority view. However, there is a wider context that may be shifting the balance in favour of some degree of harmonization. The sharing of knowledge in an increasingly global scientific community is challenging the parochialism of work that only remains accessible within the boundaries of nation-states. The social realities of the postmodern world, too, are those of increasing international migration, ethnic diasporas, the emergence of transnational identities, the adoption of multiple identities spanning gender, nationality, ethnicity, race and religion, and an emerging global civic society. The resolution of ethnic conflicts, too, have increasingly come under the jurisdiction of international bodies. A theoretical framework that, at least, explores this wider context – without eschewing the historical specificity – looks to be an increasingly needed approach. This section investigates the issue of historical specificity and the extent to which it might constrain a cross-national comparative approach. It then reviews three methods that have been attempted to advance such comparative work in the field of race/ethnicity: semantic locality, input/output harmonization and metric approaches.
In the field of ethnicity and race, the meaning of concepts and the use of terms is heavily context dependent. It has been established that an essential part of that context is the conceptual framework used to collect the data on race/ethnicity. It will determine, for example, whether the categorization is based on official constructs from the census, a set of socially modelled options or the individual subjective conceptualizations of group members. Migration histories will frequently determine the use of specific terms for groups. Because most migrants from Asia originated in the Indian subcontinent in the postwar flows to Britain, for example, the term ‘Asian’ became synonymous with migrants who had originated in South Asia. In the US, by contrast, 20th-century migration from Asia has been characterized by a much broader country of origin range, including China, Japan, Korea, Vietnam, Cambodia and the Philippines, all of whom are encompassed in the term ‘Asian Americans’. Similarly, the salient official terms for the Black population in Britain are ‘Black African’, ‘Black Caribbean’ and, recently, ‘Black British’, reflecting countries of origin and the demands of those born in Britain. In the US, the census terms are ‘Black, African American, or Negro’, the first two being the preferred terms among 44 percent and 28 percent, respectively, of the Black population (Tucker et al., 1996), ‘Negro’ – preferred by just 3 percent (mainly elderly people in the Southern states) – being a legacy of the language of slavery and plantation societies.

In some countries, the problem of context and conceptual definition may be even more layered, manifold, subtle and complex. For example, the carefully nuanced meanings of the terms ‘Pa-keha’, ‘European’ and ‘New Zealander’ may depend on such factors as source of definition and interpretation (experts, officialdom and the public), skin colour, ancestral origins, degree of cultural affiliation with country of ancestral origins, length of residence in New Zealand and even where people live (Allan, 2001). The risks of separating terminology and context may be a perceived reification of terms and an assumed solidity or concreteness of meaning. The multiple forms of discourse that have characterized this field are probably necessary for a number of reasons. Some of the underlying tentativeness of the concepts themselves gives rise to shifting terminology. Much of the discourse is deeply rooted, historically determined and specific, resulting in some of the terminology being infused with conflict and power conceptions of social reality.

It is difficult to see how these problems of context can be dealt with in the economy of an internationally applicable and agreed glossary of synthetic analytical terminology that has been recommended by epidemiologists. The range of possible contextual factors to which meaning is inextricably linked is wide (to which, of course, the glossary definitions themselves would add), as often is the distance between theoretical pronouncements
and the way concepts and terms are used in practice. Other definitive approaches to terms and their definitions – like a more robust system of terminological bookkeeping – have been taken by institutions like journal editors (Bhopal et al., 2000) and the judiciary. However, such formalization and, inevitably, reduction in the concepts and terms used are unlikely to be a desirable response to the plurality of language in the wider sphere. Since both inadequate conceptualization and the inappropriate use of labels and terms are well-recognized – and probably mutually reinforcing – problems and barriers to an improved cross-national understanding of race/ethnicity terms, a more discursive approach to concepts and terms may be required. Some have argued that functional equivalence can best be achieved through the method of semantic locality, although the empirical referents for common concepts frequently suggest that the concepts themselves are representative of a range of meanings.

**Semantic Locality**
Attempts have been made to facilitate the recognition of specific concepts and their related terms drawn from different vocabularies through the organizing principle of semantic locality, whereby concepts from different vocabularies can be understood and located by their relationships to other concepts, as exemplified by the Unified Medical Language System (UMLS) (Bodenreider et al., 2001). However, the lack of consistency in the use of ethnicity and race concepts and terminology sets limits to this method for representing relationships. In the late 1970s and 1980s, UNESCO sponsored a pilot of the INTERCONCEPT project. The Committee on Conceptual and Terminological Analysis (COCTA) of the International Social Science Council sought to analyse and solve conceptual and terminological problems in the social sciences using word-to-meaning and concept-to-terms approaches (Riggs, 1981; Sartori, 1984), resulting in some INTERCOCTA glossaries, including concepts and terms used in ethnicity research. However, such primarily US-driven approaches were not whole-heartedly endorsed by the wider academy (Higgins et al., 1979) and their legacy has been meagre in terms of lasting and usable methodologies.

**Input and Output Harmonization**
This approach has been advocated by Hoffmeyer-Zlotnik (2003) and other practitioners involved in comparative social research in Europe, possibly given the somewhat different approaches to ethnicity data collection in continental Europe compared to those in Britain and North America. In the former, there has been a much stronger focus on the dimensions of ethnicity – such as citizenship, country of birth, language, religious denomination, migrant status and nationality – and less on race or ethnicity as a unified concept representing self-identity. This has been
revealed in a recent survey of the practices of the EUROSTAT countries in their civil registration processes (Aspinall et al., 2002): most of these countries in official data collections have focused on such matters as nationality status, citizenship, country of birth and even language but to the substantial exclusion of race and ethnicity, using a variety of methods including decennial censuses, annual micro-censuses and population register-based and other administrative collections.

The multidimensional approach to measurement is less favoured in Britain and North America. For example, it is eschewed by Witzig (1996), who argues that the independent determination of every variable that makes up ethnicity, ‘thus rendering the global concept of ethnicity obsolete for the health management of individual persons’, is too simplistic. Rather, he sees ethnicity as an important part of self-identity, thus acknowledging a long tradition by critical social theorists of developing the ‘group identity’ approach to ethnicity. While the decomposition of the concept of ethnicity may be appropriate in some contexts, it is generally agreed that some measure of ethnic/racial self-identity is still needed. McKenzie (1998) has argued for a similarly cautious approach to that of Witzig (1996), contending that it is not always desirable or even possible to ‘unpack’ culture or ethnicity since they are context-driven social and psychological concepts and it is difficult to know how each factor affects the whole.

Within these qualifications, there is clearly some scope for input harmonization – in which data are collected in the same form in each country – where the focus is on relatively straightforward dimensions like respondent and parental country of birth, language use and migrant status. There are, for example, some international standards with respect to country names and language (e.g. the ISO 639 Language Codes). Output harmonization is where the selection method is left to each country but the aim is to measure an internationally agreed concept. The latter may be done ex ante, planned from the start, or ex post, when a conversion procedure is used to adapt national statistics. When the differences in national systems are such that the same (or very similar) questions cannot meaningfully be asked, input harmonization is of little value. For example, common cross-national questions on religion and ethnic/racial self-identity cannot be asked as the overarching concepts and groups or collectivities are specific to each country and not transferable.

Attempts to harmonize questions on religion have created difficulties for reason of nationally specific contexts (Erikson and Jonsson, 2004; Kolsrud and Skjåk, 2004): differences in historical developments of religious institutions across countries, different colonial and migration histories leading to different religious minorities, different secular traditions, and differences in the national structures of religious organizations, including the presence or absence of state churches. The European Social Survey (ESS),
for example, settled for a mix of both input and output harmonization for the religious variables. The same question was asked in all countries, the answer categories (religions and denominations) being set by the national teams to suit the needs of their own country. These responses were then recoded into a common ESS coding frame for religious denominations, thus bridging answers to a common standard.

This approach is equally problematic for ethnic/racial identity. Given the now widely accepted postmodern interpretation of ethnic/racial identity, as one that is constantly negotiated and renegotiated in interaction, Billiet (2004: 388) concluded that ‘the measurement of ethnic identity in surveys by means of standardised instruments seems impossible’. The ESS’s Central Coordinating Team (CCT) came to the same conclusion: ‘the task of drawing up exhaustive country-specific lists and devising a generic code frame into which these could be post-coded would have proved too difficult in the context of the ESS’ (Billiet, 2004: 414). Even if the focus is narrowed down to ethnic or family origins, the overwhelming number and diversity of such origins and their country-specific nature is likely to preclude measurement in surveys. Their national presence is likely to reflect a complex set of factors, including the colonial histories of the different nation-states, the characteristics of labour markets that characterize the experience of economic migrants and increasing flows of refugees and asylum seekers from an ever-widening pool of contributor countries. Output harmonization of answers to ethnic self-identity questions would result in long lists of idiosyncratic terms, which would be difficult to collapse on the basis of any common points of reference.

Moreover, ethnic and racial self-identity are concepts which are frequently (and increasingly) inextricably linked to national and subnational identity, further complicating measurement. In Britain, this has been reflected in the use of bicultural labels (like ‘Black Welsh’ and ‘Scottish Pakistani’) and a popular movement in Wales to have ‘Welsh’ added to the list of cultural background options for the ‘White’ ethnic group in the England and Wales 2001 Census. In response to the need for collection of national identity information, in 2003 the Office for National Statistics developed a ‘national group’ question as part of a two-question data presentation on ethnic and national group. In Canada, an increasing number of respondents are identifying as ‘Canadian’ in the question on the ethnic or cultural group(s) to which the person’s ancestors belonged: almost 40 percent of the single and multiple responses in the 2001 Census. The importance of a national identity to Canada’s mixed-race population has already been noted. National and subnational identity has long been important in many European countries, including Spain, Belgium, Latvia, Slovakia, Bulgaria and others. The fact that survey respondents do not consistently adopt either a national or subnational
identity and in some cases view them as complementary adds to the complexity.

Harmonization methods, however, may enable some measure to be taken of the ethnic mix in a country by multidimensional measurement. This would appear to be most productive when there is a focus on the separate dimensions (such as respondent/parental country of birth, nationality, language, religion, etc.) in their own right. It is frequently these separate dimensions that are of most value to policy-makers in their efforts to provide equal access to services, for example. It seems more doubtful that such an approach can serve as a proxy for ethnic/racial group measurement. An analysis of such dimensions across over 100 ethnic groups yielded at least 14 separate features (Leets et al., n.d., cited in Billiet, 2004). Again, Billiet concedes that some measure of subjective self-definition would still be needed, concluding that: ‘Going that direction would suggest a set of complex multiple indicators, and some kind of cluster analysis afterwards in order to construct the ethnic groups ad hoc by means of closeness and distance.’

Thus, attempting to identify a particular mix of attributes to ‘profile’ an ethnic group does not appear to be a defensible strategy. It is difficult to know how each of the dimensions affects the whole (or global) concept, which is social, situational and relational in character. Clearly, too, ethnic groups will draw in differing measure upon the various dimensions. For example, religion is a fundamental part of South Asian identity in Britain but perhaps less so with respect to Black identities (although a fifth of Black Africans identified as Muslim in the 2001 England and Wales Census). Parental country of birth may be of limited value in characterizing long-established communities like the Irish. Such approaches, then, cannot be a substitute for measurement of ethnic/racial self-identity, in itself, and should not be confused with such. The ESS CCT concluded that the only such question they could ask was: ‘Do you belong to a minority ethnic group in [country]?’ However, there appears to be much scope for further investigating, in a cross-national survey context, the scope for input and output harmonization of the various dimensions of ethnicity. More empirical work is needed to identify appropriate strategies, to assess how successful this approach is at capturing the level of ethnic diversity in different nation-states, and the utility of such work in contributing to a better understanding of cross-cultural differences that are relevant to public policy.

**Numerical Indexing of Different Ethnicity Categories**

Finally, Lambert (2004; see also Lambert and Penn, 2001) has investigated a method of ethnic category scoring based on the collection of categorical data in terms of the various dimensions or ‘ethnic referents’, which are
then analysed by assigning metric scores to different categorical levels within each country. He discusses several ways in which ethnic categories could reasonably be assigned score values (including simple ranking and more complex methods such as the use of stereotyped ordered regression models) and describes how this approach can present the typical effects of ethnicity differences as they operate through a given structure of inequalities. One of the potential values identified of such ethnic category scoring is the parsimony with which it summarizes the key elements of ethnic category differences that affect populations, but Lambert was forced to conclude that: ‘representation of ethnic category differences through score values still lacks convincing displays of its empirical superiority – though the possibilities remain open.’ It is important that such analytical strategies continue to be investigated as there is a demonstrable need for the cross-national comparative analysis of ethnic differences. Moreover, such problems are not atypical in the survey context: there are frequently constructs of interest that cannot be observed directly but measured only indirectly by means of observable indicators. Various types of scaling techniques have been developed for deriving information on these ‘latent variables’.

**Conclusions**

What this article has demonstrated is that conceptual bases for ethnicity and race include both social categories and group identities, each generating its own specific vocabulary. Many terms are geographically specific, including those for pan-ethnic collectivities like ‘Asian’ and ‘White’. Recognition by officialdom can have ethical implications and can lead to the exclusion of groups. All these factors and the layered, manifold and subtle contexts of use that are intricately linked to terminology favour an approach that explores the connotative reach of the different concepts and terms with respect to their context rather than one that seeks to systematize knowledge in the form of the economy of glossary. The wider agenda of how cross-national comparisons can be achieved remains largely unexplored. International agencies like the United Nations and EUROSTAT report that they have no internationally recognized standards or classifications for ethnicity or related concepts and that no such criteria can be recommended. Given the specificity and distinctiveness of each country’s individual processes of ethnogenesis, in which context terminology has evolved and developed through sociohistorical and political processes frequently implicating the local state, two concurrent approaches may yield results.

First, an initial documentation project is much needed, involving a country-by-country analysis prepared to a standard format. It could
include exploration of the prevalent conceptual terms that are known to vary in usage across countries (like race, population group, ethnic group, ethnic/family origins, ancestry, national origins, citizenship, nationality, etc.), the use of official terminology, the popular conceptualizations of race/ethnicity, and the collective and specific group identifiers used by community members. This would assist interpretation as the definition of concepts and terms offered by officialdom frequently differs from those within the academy and the vocabulary used by members of the wider society. The work of Stephens with respect to the countries of the Caribbean and South America (Stephens, 1999, 2003) offers an exemplary country/regional analysis. Second, this exercise, of itself, would then enhance the scope for greater cross-national comparative work in the field of race/ethnicity as there would be a resource to facilitate interpretation of terminology that is embedded in national-specific contexts. It may also facilitate various means of mapping specific groups cross-nationally based on selective criteria, such as country of origin. This kind of sociology of knowledge approach is likely to avoid undue abstraction in discussing differences and, by revealing the complexity of the terms and the way they have developed, may contribute to a greater understanding of the rhetoric involved and even the development of theoretical ideas in this area.

Efforts focused on harmonization for global variables like ethnic identity and dimensions such as religion have proved problematic and, in some cases, self-limiting. Other similar initiatives – such as the thesauri used for medical indexing – are extensively criticized with respect to their shallow structures and inappropriate language for ethnic/racial terminology (Efthimiadis and Afifi, 1996). However, it is likely that substantially more progress can be made with the dimensions of ethnicity, such as respondent and parental country of birth, language and migration status and years since migration. A promising start in the various initiatives to harmonize social variables across the European suite of social surveys needs to be expanded to wider cross-national contexts. Issues of the meaning of ethnic/racial concepts and terms in such a comparative context remain of crucial importance to political institutions and scientific communities. Given the currently unprecedented interest in these issues, as witnessed by activity generated by the new census classifications, the development of national policies and international collaborations to ameliorate the underlying causes of racial and ethnic disparities, the increasing integration of the work of the statistical offices and other institutions of the EU, and recent revisions of the controlled vocabularies of health sciences databases, it is an opportune time to support and promote these harmonization objectives.
References


Biographical Note: Peter Aspinall is a senior research fellow in the School of Social Policy, Sociology and Social Research at the University of Kent, UK. His research interests are the conceptualization of ethnicity and race, the operationalization of these concepts in official contexts like national population censuses, and ethnic disparities in health and health care. He is currently honorary special advisor to the London Health Observatory and has recently completed a secondment to the Department of Health’s Equality and Human Rights Group. Between 1994 and 1999 he was National Convenor of the Office for National Statistics’ Working Subgroup on Ethnicity for the 2001 Census Development Programme. Peter Aspinall has published extensively in the field of ethnicity in Sociology of Health and Illness, Social Science and Medicine, Sociology, Policy and Politics, Social Policy and Administration, British Medical Journal, Journal of Public Health and others.

Address: Centre for Health Services Studies, School of Social Policy, Sociology, and Social Research, University of Kent, George Allen Wing, Canterbury, Kent CT2 7NF, UK. [email: P.J.Aspinall@kent.ac.uk]