Challenges of Measuring an Ethnic World
Science, politics and reality

Proceedings of the Joint Canada-United States Conference on the Measurement of Ethnicity
April 1 - 3, 1992

Issued September 1993

Minister responsible for Statistics Canada
Jean J. Charest

STATISTICS CANADA
Ivan P. Fellegi, Chief Statistician of Canada

U.S. Department of Commerce
Ronald H. Brown, Secretary
Economics and Statistics Administration
BUREAU OF THE CENSUS
Harry A. Scarr, Acting Director

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-042049-0
SUGGESTED CITATION

Statistics Canada and U.S. Bureau of the Census, Challenges of Measuring an Ethnic World: Science, politics and reality

Canadian Cataloguing in Publication Data

Challenges of measuring an ethnic world : science, politics and reality : proceedings of the Joint Canada-United States Conference on the Measurement of Ethnicity, April 1-3, 1992

Text in English and French with French text on inverted pages.
Title on added t.p.: Les défis que pose la mesure de l'origine ethnique : science, politique et réalité : Conférence canado-américaine sur la mesure de l'origine ethnique le 3 avril 1992.


GN495.6 C42 1993 305.8/0021 C93-099397-7E
How to Measure Ethnicity: An Immodest Proposal

Charles Hirschman
University of Washington

A standard theory of measurement is that practice makes perfect or at least that standardized measurement systems tend to reduce error. This policy is one of the linchpins of social statistics gathered in national systems of population censuses and surveys. While experienced users of social statistics are acutely aware of the remaining errors in the data, there is no doubt that the standardization of questions and survey methods, the professional training of interviewers and coders and careful evaluation of all phases of survey and census procedures have improved the quality of data available to social science and policy making.

For the area of ethnicity, however, statisticians and others concerned with data collection have begun to doubt the conventional strategies of measurement. Attempts to improve the measurement of ethnic categories in censuses and surveys have been frustrated by the seeming inability of respondents to give consistent or meaningful responses. The examples of "contradictory" responses for measures of race, ethnicity, ancestry, language, birthplace and similar questions are familiar to every statistician and scholar (Levin and Farley 1982; Lieberson and Santi 1985, Khoo 1991). For the censuses of the United States, the addition of new questions on Hispanic origin and ancestry has created more data but has given rise to even more uncertainty over the ethnic composition of the population and raised new questions on the meaning of ethnic identity (Farley 1991).

For statisticians and researchers who rely on hard statistical data, the mushy data on ethnicity have been the source of great consternation. The first reaction is to fix the data by measuring errors (inconsistent responses) and making appropriate adjustments and then to create better questions that elicit "truer" or more consistent responses. The problem is that there is no one method to establish validity of responses. Is the true number of American Indians (native Americans) best measured by those who claim American Indian identity on the race question or those who report American Indian ancestry (Snipp 1989, 47-61)? It is possible to "explain" the differences in response to such questions (perhaps the race question best identifies those with full or almost full American Indian ancestry, while the ancestry question also includes those with partial American Indian ancestry) but it is unlikely that all persons weighed "objective" criteria in the same way or even that individuals report their own ethnicity consistently in different data sources. Probing surveys of ethnic identity and related attitudes and behavior have found only weak ethnic attachments for the majority of Americans (Alba 1990). Waters (1990) reports that many White Americans choose their ethnicity among a variety of options for quite idiosyncratic reasons.
Because of the lack of "objectivity" of ethnicity, another common reaction of statisticians is to simply claim that reliable data cannot be measured on the topic of ethnicity. There are many important questions that surveys and censuses measure poorly and perhaps ethnicity is one of them. The problem is that the public and policy makers do not accept this option. In many if not all societies, ethnicity appears to be relatively unambiguous with numerous physical and cultural attributes differentiating the population. For individuals who are sure of their own ethnicity and that of others, the ambiguity of ethnicity seems to be a minor problem. Moreover, in many societies, public policies are formulated on the basis of the size and composition of ethnic communities. With considerable budgetary and political factors hanging in the balance, decision makers are reluctant to be told that it is impossible to measure ethnicity.

The motivation for this conference and of my paper is to rethink the concept of ethnicity in hopes of designing improved strategies of measurement. I begin with an historical review of the concept of ethnicity and then review the problems of reliable measurement. The final section of the paper presents some recommendations for a more consistent approach to measure ethnicity in censuses and surveys.

The Content of Ethnicity

What is ethnicity? The fact that there is no simple answer to this question illustrates the depth of the problem before us (Glazer and Moynihan 1975; Petersen, Novak, and Gleason 1982). The question of deciding who is a Jew has sharply divided religious leaders, public officials and scholars in Israel for many years. In some societies, the question of whether a specific person is a member of a particular ethnic community has become a matter for judges and juries to decide. While it might seem that ethnic identity should be entirely a personal matter, the issue of classification can become contentious if there are state-sponsored entitlements such as scholarships or employment that are available to some groups and not others. The same problem arises when discriminators are trying to figure out who should be barred from a country club, a fraternity or employment. In many cases, elaborate investigations usually produce contradictory evidence, arguments are made on the basis of hair-splitting logic, and ad hoc judgments prevail.

In Malaysia, the constitution has codified the definition of ethnicity. According to the constitution, a Malay (the indigenous population of the country) is a person who professes the Muslim religion, habitually speaks the Malay language, conforms to the Malay custom and was born in the country before Independence (or is the issue of such a person). While such a formal definition may seem to be the ultimate step in maintaining rigid ethnic divisions, the reality is much more fuzzy. The authoritative legal text on the Malaysian constitution written by the former Lord President (Chief Justice) of the Federal Court writes that to be a Malay for the purpose of the constitution you need not be of Malay ethnic origin (Suffian bin Hashim 1976, 291). Suffian explicitly acknowledges that acculturation could be the basis of becoming a
Malay. Indeed, this is consistent with the traditional definition that conversion to Islam was all that was necessary to become accepted as part of the Malay community.

The Assumption of Mutually Exclusive Ethnic Groups

The standard assumption of ethnicity is that there are visible differences among populations, defined in terms of phenotype or culture, that can be used to identify all individuals into a finite number of mutually exclusive ethnic groups. Ethnic differences are generally assumed to have arisen in prehistorical times when geographical isolation across many generations allowed for natural variations to develop. There are, however, several flaws to this argument.

Although geographical isolation in prehistory did lead to extensive physical and cultural variations, patterns of long distance migration meant that there were also opportunities for contact and intermarriage among different populations (Davis 1974). The result is that present-day populations that claim a distinct identity and ancestry may well be hybrid populations. For some populations that originated in the modern era (e.g., the English and many New World populations) the fusion of different physical and cultural stocks is well known but similar processes may have occurred in prehistory for many (most) other groups. Physical anthropologists have concluded that there is only a modest degree of genetic variability (relative to total variability) between populations (Chakraborty 1986, 35; Polednak 1989, 32).

Over the last few hundred years, even as ethnic divisions were becoming synonymous with the identity of modern states and caste-like "race relations" were being constructed on a world scale, there has been widespread ethnic blending (Hirschman 1991). By ethnic blending, I refer to inter-ethnic unions (interbreeding) and shifts in ethnic affiliation. In general, however, much of the evidence of ethnic blending is obscured because history is selectively revised to fit present day perceptions. The contemporary presence of states with a core ethnic community is taken as evidence for the historical origins and continuity of the ethnic group. Even most minority groups typically claim historical continuity as endogamous communities.

There is considerable evidence, however, that most ethnic communities are either amalgams of different peoples or have absorbed significant numbers of other groups through conquest, the expansion of national boundaries and acculturation. The creation of slave societies in the New World and other long-distance labor migration systems over the last 500 years have created some of the most obvious examples of blended populations. Although there are tremendous variations across countries in South America, North America and the Caribbean, there has been widespread admixtures of Native American, African and European populations throughout the New World. National systems of ethnic divisions and classification schemes seem more related to political history than to ancestry or cultural divisions (Harris 1964; Van den Berghe 1967).
How to Measure Ethnicity: An Immodest Proposal

Numerous examples of blended populations can be found any place on the globe. The creation of the modern French and Italian nations are examples of how language acquisition and national identity can change within a few generations. In mainland Southeast Asia, I have argued that the cultural core of the Vietnamese, Thai and Burmese populations has grown through successive waves of absorption of different ethnic-linguistic groups (Hirschman 1991). The construction of Han identity in modern China may have similarly been an historical product of the large-scale cultural absorption of disparate ethnic groups.

The Problem of Permeable Boundaries

A common belief is that the major problem of consistent ethnic measurement is the classification of peoples on the margins, e.g., the children of mixed marriages, individuals with characteristics of different groups or people who claim to have no ethnic identity. In many countries, the number of inconsistent cases is generally small and thought to be of recent origin, so the hope has been to resolve "problem cases" with ad hoc "rules of thumb". For example, a person of mixed ancestry could be classified according to the ethnicity of his father (or mother) or according to his primary ethnic identity or even coded with multiple ethnic affiliations. The hope is to maintain the objective basis of ethnicity but with a small degree of reliance on subjective criteria for those at the margins.

The preceding review, however, raises serious doubts that any assumptions for an objective base of ethnic identity can be justified. While the number of marginal cases may be a small fraction of cases at any one moment (or can be made small by framing broader categories), the reality is that many individuals in most societies could be classified in multiple categories if additional criteria were used or if respondents had complete information on their ancestry. The contemporary evidence of permeable boundaries reflects only the ambiguity based on the most recent generations. A historical view yields a kaleidoscope of shifting ethnic affiliations that have varied considerably in both premodern and modern times. Moreover, there is a fair degree of ethnographic evidence that many individuals may have multiple ethnic identities that are "put on" according to the situation or context (Leach 1954; Lehman 1967; Nagata 1974).

The Importance of Ethnicity

If the previous account is only partially true, then there is little objective basis for the ethnic classifications used in most societies. Moreover, there is a significant degree of unavoidable inconsistency in the measurement of individual ethnic affiliation. This interpretation is consistent with the evidence that variations in the format of questions on ethnicity (open versus closed ended) and variations in the number of categories listed as choices or examples elicit such different responses (Pryor et al. 1992).
How to Measure Ethnicity: An Immodest Proposal

Intuitively, the ambiguity of ethnicity classifications would seem to minimize the importance of ethnicity in the modern world. Yet exactly the opposite conclusion seems to be evident from even a casual reading of twentieth-century history and contemporary current events. The ideological prop for the colonial rule throughout much of Asia and Africa for the first half of the century was largely a racial construction. The apartheid structure of modern South Africa, Jim Crow in the United States, and Nazi Germany are only the most odious examples of the widespread racism that held sway throughout the Western world for much of this century. And even if one might claim that these practices are losing ground, it is clear that the ethnic passions still dominate political life in Northern Ireland, the former Yugoslavia and other parts of eastern Europe. Even without violence, ethnic tension is probably the central political issue in the United States, Canada, India, Malaysia and many other societies around the globe.

The reconciliation between ethnicity as the defining issue in social life and ethnicity as a concept so ambiguous that it is difficult to measure seems to be an impossible task. I believe that the key to the problem is the distinction between the study of long-term social dynamics and short-term political struggles. Over the long term there is a fair degree of mobility across ethnic divisions and even wholesale changes in the structure of ethnic classifications. The criteria used to create ethnic divisions are largely arbitrary and many of them can be manipulated by motivated individuals. Even for the ethnic characteristics that are inherited there is a moderate degree of variation that makes many individuals difficult to classify unambiguously. In the short run, ethnic divisions in many societies are strongly correlated with social, economic and political status. State policies are often shaped by groups that wish to reinforce the advantages of their own group and/or maintain the subordination of other groups. Not too surprisingly, ethnic groups, like families, often discriminate in favor of their own. These practices by states, families and individuals are what fuels the fires of ethnic antagonisms and makes ethnic disputes among the most volatile of social problems.

An analogy can be made between class relations and ethnic relations. Over time, especially over generations, there can be a fair degree of social mobility and even realignment of the class structure. But these long-term processes do not preclude strong class antagonisms resulting from the breakdown of patron-client structures, industrial strikes, state policies or other political and economic events. A high degree of social mobility probably tempers class antagonisms because family relationships can cross-cut class divisions. The same principle probably holds for ethnic relations but we really do not know the parameters of such relationships. The key conclusion is that contemporary ethnic passions and conflicts do not inevitably mean that there has been historical constancy of ethnic definitions or boundaries.

The Creation of Ethnic Classifications

Definitions of ethnic groups might be illustrated with an analogy to the scope of traditional polities. In past times, empires and other political units had a clear center and heartland but
only vague and shifting frontiers. As power waxed, more people and territory were included under the sway of the central state. At other times, those in the outlying provinces had only loose connections to the traditional center and may feel compelled to re-identify with expanding political units from other areas. In an analogous fashion, definitions of ethnic groups are likely to describe the physical and cultural features of the core constituencies but may not accurately fit those at the margins or boundaries.

This lack of clear boundaries makes it very difficult to develop consistent measurements of ethnic groups. Any definition of an ethnic group will capture the core community but the inclusion of those at the boundaries will depend on the nature of the question and the alternative choices. The standard practice for the measurement of ethnicity in censuses and surveys, as with most other social categories, is to follow precedent. The designers of census and survey questionnaires examine earlier classification schemes and typically make incremental adjustments from the prevailing standards. Continuity of measurement is highly valued because it allows for systematic comparisons across time. Continuity also reinforces beliefs that social measurements reflect real distinctions.

The "race" question has been part of the censuses of the United States for 200 years. The categories have changed and clearly the contemporary understanding of race is different than what it once was. It is difficult to provide a convincing conceptual justification for the current race question beyond the fact that it provides comparability with past measures (Lieberson and Waters 1988, 15). The question may have other advantages such as a mutually exclusive set of categories but these are independent of its conceptual base.

The idea of ethnic classifications as a set of arbitrary, but not accidental, categories is reinforced by studies of the historical evolution of ethnic classifications in the censuses of Hawaii (Petersen 1969) and Malaysia (Hirschman 1987). In Hawaii, Petersen (1969) found that changes in census ethnic classifications were more related to political factors than to demographic considerations. The position of an ethnic community in the local power or prestige structure was much more important than its population size in determining the timing of inclusion as an ethnic category in the census. Over time one could read the changing political dynamics of Hawaii from changes in census classifications.

In my analysis of the evolution of the measurement of ethnicity in Malaysia from the first census of the Straits Settlements in 1871 through the colonial era and the postindependence period, the impact of political and ideological factors was dominant. In the earliest classifications, categories were listed alphabetically and referred to as nationalities. By the turn of the century, the categories were termed as races and there were many signs that the Social Darwinian conception of races had taken hold (Hirschman 1986). Europeans were always listed at the top of every table and the classification of local ethnic groups reflected European racial thinking.
As political independence approached, there were changes in the composition and order of census ethnic classifications.

In general I conclude that official ethnic classifications tell as much about the society in question and its ethnic political balance in particular than the physical or cultural distinctions among peoples. In any multiethnic society there are a variety of alternative classifications that can be used to classify human ethnic diversity. Different measurement strategies produce different and often inconsistent results. Since following precedent is not an unbiased strategy and there is no "truth" or validity check, it is not surprising that statisticians are turning to sociologists and anthropologists for advice.

How to Measure Ethnicity Better

After painting this rather bleak picture of the problems of measuring ethnicity, it is only fair to ask how I think it might be done better. At the outset I must repeat the obvious: there is no magic bullet. Ethnicity is a multidimensional phenomena with both phenotypic and cultural dimensions. Individuals may have multiple ethnic identities with varied levels of attachment and these identities may or may not be correlated with any objective characteristics. In spite of these problems I think it is possible to create questionnaire items that will have conceptual clarity and reliable measurement.

There are two dimensions of ethnicity that stand out above all others. The first is an individual’s primary identity among the major ethnic groups in a society. The second is an indicator of an individual’s descent or ancestry from among a broad range of ethnic groups. Let me explain the rationale behind each of these concepts and illustrate them with some sample questions.

Primary Ethnic Identity

The first dimension addresses the question of why ethnicity is important. According to the prior discussion, ethnicity is important when it has instrumental value (or disvalue) through an association with political, social or economic factors. If the government and other institutions, social groups or individuals use ethnic criteria to provide rewards or access to scarce resources and/or select individuals for participation in formal or informal associations, then ethnicity matters.

This dimension is logically measured only in terms of mutually exclusive assignment among a predetermined list of major ethnic groups. While individuals may have multiple ethnic identities, it is unlikely that policies of inclusion or exclusion can rest on the ambiguity of joint ethnic membership. Consider an American who has one Japanese parent and one White parent. In terms of the issue emphasized there, the question is how the person’s life chances are affected by his/her own primary ethnic identity and/or others’ perceptions. While the survey or census
cannot measure others’ perceptions, the respondent’s primary ethnic identity is probably influenced by how s/he is seen by others. Because the politics of ethnicity in American society is typically based on mutually exclusive membership, I do not see a joint ethnic status as a meaningful category for instrumental purposes. In other societies, however, blended categories such as Eurasian or Mestizo may have a recognized status, both in terms of community identity and in societal perceptions.

The question could be phased as follows:

AMONG THE LIST OF GROUPS LISTED BELOW (OR ON A CARD HANDED TO THE RESPONDENT), WHICH DO YOU CONSIDER THE ONE WHICH IS CLOSEST TO YOUR PRIMARY IDENTITY?

If the respondent claims to not have an ethnic attachment or multiple attachments, the follow-up might be phrased:

YES, I UNDERSTAND. BUT IS THERE ONE GROUP THAT MIGHT BEST DESCRIBE HOW YOU ARE SEEN BY OTHERS?

The selection of groups to be included on the list of choices is the critical aspect of this question. The concept of major groups means that the selection must be limited to groups that are large enough to be visible. Visibility is not based on physical or cultural distinctiveness but on demographic and political criteria. A major group must have some corporate image of themselves and/or been seen by others as a distinct community. For example, the term Asian and Pacific Islander (a U.S. census category) is meaningless because no—one not the groups included nor anyone else—thinks or acts in reference to such a diffuse label. The other key attribute is a minimum population size. Consider, for example, an American who reports his/her primary ethnic identity to be Armenian. Armenian is a perfectly appropriate ethnic category with a distinct historical identity. But the membership in this category is too small to be considered politically important in the United States and thus be subject to preferential and/or discriminatory treatment.

The list of major ethnic categories will vary in different societies and over time in the same society (comparability should be maintained by aggregation of categories). The construction of the list of categories should be based on extensive research using focus groups to select potential categories and then trial surveys to check on recognition of categories. In general, the effort should be to minimize the total number of categories (perhaps excluding those with less than one percent identification of the population) and also to minimize the numbers who claim no primary ethnic identification. This will involve tradeoffs and there may not be a single optimal list of politically important ethnic groups.
Ancestry

The aim behind this concept is to measure the diversity of the population by asking the national or ethnic origins of their ancestors. For some people their primary ethnic identity and ancestry will be the same. But for many others ancestry will reveal varied ethnic roots which may or may not be related to current identification, cultural attributes or physical features. The primary measurement problem will be that most individuals will not know the true ethnic origins of their ancestry beyond their parents and grandparents. However, the question will provide a minimal estimate of the diversity of the ethnic origins of a population.

In addition to being of great public interest, ancestry data can help to clarify the concepts of race and ethnicity. The popular assumption is one of homogeneity of ethnic identity, cultural attributes and other characteristics. However, most research has shown that these relationships are only loosely connected and are subject to considerable flux. It is important to obtain a measure of the complexity of ethnic origins and then to measure the overlap of ancestry, ethnic identity and other ethnic criteria (language use, birthplace, cultural attributes).

In order to spread the net as broadly as possible, the suggested question might be:

THINKING ABOUT YOUR PARENTS, YOUR GRANDPARENTS, AND YOUR ANCESTORS, WHAT NATIONALITIES OR ETHNIC GROUPS ARE REPRESENTED IN YOUR FAMILY’S HISTORY? [The respondent could be prodded with a list of nationality and ethnic groups.]

There should not be any limit on the number of responses that a person should give. It would be preferable if the respondent could rank-order the responses. While this might be possible in a survey, I understand that the time needed to rank-order responses would probably be prohibitive in a population census. An additional problem would be coding a variable number of responses.

Conclusion

Census measures of race and ethnicity originated in an earlier era when assumptions about the biological basis of racial distinctions and firm links between cultural characteristics and ethnicity were taken for granted. Because of revised thinking about these assumptions and the need to provide broader measures of ethnicity, new questions have been added to the United States census (and those in other countries) to identify groups on the basis of identification and ancestry. However, these new measures have, in general, been a disappointment to those who manage the nation’s social statistics. The new measures (and the old ones) seem to be full of errors and unexpected complexity. Individuals provide answers that are inconsistent and responses seem to vary enormously with slight changes in the structure of the questions.
How to Measure Ethnicity: An Immodest Proposal

Given this predicament, it seems wise to rethink the concept of ethnicity and review the base of empirical knowledge. My conclusion, after reviewing the literature, is that there is little objective basis for the conventional model of ethnic groups as endogamous populations with distinct cultural or phenotypic characteristics. Extensive patterns of ethnic blending in prehistory and in the modern era mean there is substantial overlap in the ethnic origins (and identities) in almost every population. Moreover, social and cultural change over the last century has resulted in very weak ethnic attachments for many persons in multietnic societies. These processes are at the root of the problems of measurement of ethnicity.

This does not mean that ethnicity is unimportant in the modern world. Indeed, there is considerable evidence that ethnic divisions continue to be closely associated with political and economic struggles in many countries. Moreover, ethnicity is often a primary base of formal and informal social organization in many multietnic societies. In spite of the vague boundaries and the overlap of ancestry, there are still core constituencies of many ethnic groups. These groups pursue entitlements from economic and political institutions, struggle for the elimination of discrimination and organize internally to maintain solidarity. Many individuals from these groups seek neighbors, employees, friends and spouses from the pool of co-ethnics.

Given this account of the role of ethnicity in modern societies, any thoughtful strategy for measurement of ethnic attachment must set conceptual priorities. While most questions will allow identification of the core constituency of many groups (particularly for isolated groups), the ethnic composition of those on the margins and those with multiple ancestry will depend on the nature and format of the question. From my vantage point, I see two dimensions of ethnicity as being more important and more measurable than others.

The first, PRIMARY ETHNICITY, is identification with one of the major ethnic groups in a society. Major ethnic groups are defined as groups with a sufficient demographic and political presence to affect a person’s life chances. By political factors, I refer to institutional or community practices that assist, retard, include or exclude members of a group. Since such practices are group-specific, a person could only have one primary ethnic attachment. The second dimension, ANCESTRY, refers to the potential diversity of national or ethnic origins of individuals. A person could claim multiple ancestries or none.

What is left out of these concepts is important to acknowledge. Perhaps most important is the omission of any reference to the cultural base of ethnicity. The emphasis here, especially in the first dimension, is on the instrumental aspect of ethnicity. The cultural content of ethnicity may be very important, especially in the maintenance of ethnic solidarity. But ethnic groups can persist without a distinct cultural base or at least with a culture that is very similar to that of other populations. Moreover, censuses and national surveys would not seem to be the best method to collect data on cultural patterns and values. The dimensions here also leave out several of the key objective aspects often used to identify ethnic groups: birthplace, birthplace
of parents, language of origin and language used at home. Rather than use these criteria to define ethnic groups, it seems preferable to measure the association of these characteristics with ethnicity. This approach would allow for an assessment of the differences between ethnic groups to be separate from the identification of ethnic groups.

A final word on race and ethnicity is necessary. Clearly the idea of race as groups defined on the basis of physical attributes cannot be ignored. While names, dress and even language can be modified, it is all but impossible to change skin color and other physical features that affect perceptions of ethnic identity. However, I am not convinced that censuses should continue to rely on "race" questions which mix ethnic identity with assumptions about physical attributes. If we need data on physical features, it might be useful to try to measure these attributes directly and maintain the concept of ethnicity on the subjective dimensions of identity and ancestry.
References


