

**Survey2000**  
**Users' Guide and Codebook**

James C. Witte (Principal Investigator), Allison Britt, and Steve Lee  
Department of Sociology  
Clemson University

Lisa M. Amoroso, Philip E. N. Howard and Nathan Wright  
Department of Sociology  
Northwestern University

Tuesday, November 16, 1999

*Abstract*

*The Survey2000 Project is the largest and most comprehensive Internet-based social science survey to date. Along with generating interesting data about geographic mobility, feelings of community, and culinary, literary and musical tastes, the experience of operating a survey with Internet tools has set into sharp relief important methodological issues of sample size, representation, and generalization. We argue that Internet-based survey research can yield meaningfully comparable data about both Internet users and larger populations.*

KEYWORD LIST: Survey2000; Web-Based Survey Design; Sampling

Survey 2000 is a collaborative effort of the staff at National Geographic Interactive and academic researchers. Participants at the National Geographic include Valerie May, Karen Hembrough, Jennifer Kirkpatrick and Allan Mazzan. Academic contributors include William Bainbridge (National Science Foundation), Warren Belasco (University of Maryland-Baltimore County), Amy Bruckman (Georgia Institute of Technology), Bonnie Erickson (University of Toronto), Keith Hampton (University of Toronto), Wendy Griswold (Northwestern University), Carl Haub (Population Reference Bureau) Mac Parks (University of Washington), Richard Peterson (Vanderbilt University), Barry Wellman (University of Toronto), and James Witte (Clemson University). Survey and database programming was provided by Joe Germuska (Northwestern University).

# Survey2000 Users' Guide and Codebook

## Outline

- I Introduction
- II Project Development
- III Survey Content
  - A Canadian and US Adult Survey
  - B International Respondent Survey
  - C Youth Survey
- IV The Web-Based Survey Design
  - A Program Features
  - B Design Elements
- V Sample Overview
  - A The Demographics of Participation
  - B Attrition
- VI Survey logic and skip patterns
  - A Adult Survey
  - B Youth Survey
- VII Data file structure
  - A Canadian and US Adult Survey
    - 1 Demographic information
    - 2 Geographic data
    - 3 Community and cultural items
  - B International Respondent Survey
    - 1 Demographic information
    - 2 Geographic data
    - 3 Community and cultural items
  - C Youth Survey
    - 1 Demographic information
    - 2 Geographic data
    - 3 Community and cultural items

## **I Introduction**

During two months in 1998, over 80,000 people collectively spent over 2 million minutes on-line at the National Geographic Society's official web site participating in an interactive first-of-its-kind survey on mobility, community, and cultural identity.<sup>1</sup> In this paper, we describe the project background, project development, and details of the survey content and technical design. This paper also addresses the critical methodological issues of sample representativeness and bias that arise with a voluntaristic Internet-based survey.

Survey2000 was on-line at [www.nationalgeographic.com](http://www.nationalgeographic.com), the official web site of the National Geographic Society (NGS), during September and November of 1998. This project represents an unprecedented effort to use the rapidly growing power of the web to collect serious social science data. Survey2000 is a collaborative research project of the staff at National Geographic Interactive and academic researchers, funded by the National Geographic Society and Northwestern University. The survey focuses on geographic mobility, community, and cultural identity. A public use version of the data set will be available on CD-ROM to allow the academic and scientific community at large access to the data. Any information that could potentially compromise the anonymity of respondents has been eliminated from the public use data set.

Different survey instruments are used for adult Canadian and US respondents, adult respondents from other countries and for children under the age of sixteen. Data collected in Survey2000 falls into several clusters: a) migration histories, b) measures of sense of community and c) measures of cultural values and tastes in food, music, and literature. This data can be used to address important research questions including:

- 1) How have the lifetime migration histories of individuals changed over time? Large-scale surveys have looked at whether individuals have moved over the past five or ten years. Other studies (most notably by the National Center for Education Statistics and the Bureau of Labor Statistics) have looked at lifetime mobility histories for particular cohorts. Survey2000 allows researchers to study changes in migration cohorts across generations.
- 2) Where do individuals in today's society find a sense of community? Geographic communities, extended family, voluntary associations, workplace relations, telephone and computer connections are all part of the mix. But how are these different components blended and how does the balance change according to individual characteristics?
- 3) To what degree do individuals in the US, but also around the world, share common values and tastes in food, music, and literature. How are these tastes associated with demographic characteristics, including the extent to which an individual has been geographically mobile? To what extent have regional tastes in music and literature declined only to be replaced by a modulated and standardized "McCulture"?

Survey2000 also represents an experiment in survey methodology, and experiment in an area marked by great potential but also little experience (Blank, 1997; Schaefer and Dillman, 1998; Fisher et al., 1996). From this experiment we learn:

- 4) To what degree can a web-based survey replace traditional survey research methods? How successful can different methods be in addressing the sampling issue? Also, can broad-based promotion and outreach efforts extend survey coverage to the general population?

- 5) Previous experiments with computer-assisted, personal or telephone interviewing (CAPI or CATI) systems are known to have improved data quality while allowing for more complex, individually tailored interview schedules. Survey2000 builds on this strength, but then adds the interactive potential of a hyper-media interface. The goal was to produce an instrument that is not only complex and customized, but also engaging so as to minimize respondent burden and respondent attrition.

## II Project Development

Eighteen months of collaboration took place before the Survey2000 instruments began to collect data. The original impetus for the project came from staff at National Geographic Online with the idea that a web survey on the topic of population and migration would add to the NGS's coverage of modern society and millennial transition. From a survey research perspective, the nonrandom nature of a web survey sample raises serious questions; however this issues is not unique to web-based survey research and is likely to decline in significance as the web penetrates further into society (Smith 1997). The response to this challenge is detailed below in the section on sampling issues. In short, our solution depends on two mechanisms: 1) the survey relies on items from existing surveys conducted with traditional sampling and survey methods (e.g., the General Social Survey) to ensure that there are external benchmarks to assess the nature of the survey bias and construct the necessary weights, 2) extensive use of NGS public relations and community outreach resources to extend survey coverage.

The project decided to focus on geographic mobility as an independent variable and on individual values and preferences as dependent variables. Specifics of the survey instrument, however, were worked out by project collaborators. As the number of researchers interested in the data increased, so too did the commitment of the NGS. Indeed, the efforts of the academic collaborators and NGS were necessary complements to one another: academic interest in the project was necessary to build NGS support, while this support was needed to ensure the quality of the data, particularly with regard to the size and breadth of the sample.

## III Survey Content

As noted above, Survey2000 consists of three main instruments: 1) the Canadian and US Adult Survey, 2) the Youth Survey, and 3) the International Respondent Survey. Upon connecting to the Survey2000 site, respondents are asked to choose the appropriate form. As an incentive for completion, respondents are also told that a random number of participants will be awarded a gift from NGS. A follow-up screen queries each respondent's age and current citizenship; respondents who have mistakenly chosen the wrong survey form are reassigned based on this information. The major sections and key questions of each instrument are discussed below.

### III.A Canadian and US Adult Survey

The Canadian and US Adult Survey is the most complex and detailed of the three Survey2000 instruments. Respondents begin by supplying basic **demographic information**, including current primary residence, zip or postal code, marital status and household

composition. Respondents are also asked to identify languages regularly spoken in the household; however, the survey is only presented in English. Further questions ask about race and ethnicity, educational enrollment and attainment and current employment status. Separate response codes are offered to US and Canadian respondents. Questions concerning race and ethnicity are worded to prompt the respondents to self-identify as they normally do on government forms. During the survey development phase we discussed providing a greater range of response categories (including open-ended fields as used in the 1980 US census). However, we decided that the benefits of comparability with external benchmarks were greater than the richness offered by a wider range of response categories. Open-ended responses are solicited regarding current occupation and most recent occupation for those persons not currently employed.

Questions concerning **Internet access and use** constitute a second important survey section. Respondents are asked where they are completing the survey (home, work, community center, or library) and how long they have used the Internet.<sup>ii</sup> Respondents are also queried as to the frequency with which they engage in specific Internet activities (with email, purchasing products, use of listservs).

The next block of questions concerns the respondents' individual **mobility history**. Respondents are asked if they have ever lived outside the US and Canada, how long they have lived at their current address and whether or not they have always lived within thirty miles (50 kilometers) of their current address. Subsequent questions ask about the number of different dwelling units occupied, about other members of their current household, and whether or not other relatives currently live in the immediate area.

The mobility history questions then solicit respondents' place of birth. First, respondents are asked to pick from a list of geographic landmasses (North America, South America, Europe, Asia, Africa, and Oceania), followed by a list of countries associated with each. Respondents who were born in the US or Canada are then sent to a further screen that asks the respondent to choose the state or province of their birth. Based on this choice, the respondent is asked to select the closest location from a list of ten to fifteen cities in that state or province. At this point respondents are asked a similar sequence of questions regarding their residence at ages seven, fourteen, twenty-one, twenty-eight, thirty-five, forty-two, fifty, sixty, seventy, eighty and ninety. To shorten this block of questions, a filter question first determines at each age, if there has been no change in residence since the previous age. Previous responses regarding duration at current location and birth year are also used to minimize respondent burden and to avoid asking for information that can be obtained from previous replies. Internal checks helped verify that respondents were consistent in the information they provided. When respondents were inconsistent, they were asked for clarification.

Information detailing each respondent's **social world** is collected during the next block of questions. Respondents are asked how often and in what way (i.e., personal visits, phone calls, faxes, letters, cards or email) they have social contact with 1) relatives who live less than 30 miles away? 2) friends who live less than 30 miles away? 3) relatives who live more than 30 miles away? and 4) friends who live more than 30 miles away? A separate question then asks for the frequency with which the respondent gives or receives help or assistance from these same four sets of relatives and friends. Next, building on a series of questions regularly included in the General Social Survey (GSS) respondents are asked about their membership in a set of formal organizations (e.g., service clubs, veterans groups, labor unions and social advocacy groups). In addition, respondents are asked about various forms of political involvement (e.g.,

social groups). This section then closes with a series of Likert-scale (seven points ranging between strongly agree and strongly disagree) items related to community. These items<sup>iii</sup> pertain to traditional sources of community as well as Internet-based communities.

This section concludes with a series of questions about the recreational and leisure time activities that are an important part of the context of an individual's social world. Survey2000 builds on the GSS questions about involvement in a range of activities, such as gardening, reading, sports and adds additional categories, (e.g., renting videos, going to casinos and attending work-related social events) that round out the list. This section concludes with two items that assess the extent of knowledge and interest a respondent has in music, literature and food, the three areas of cultural identity that form the basis for the concluding section of the survey.

The final section of Survey2000 is titled **interests and perspectives**. Each respondent is presented with one of four randomly selected topical modules. The four topical modules are literature, food, music, and views on the world. The literature and food modules are set up quite similarly. In each module a customized list of items—authors in the case of literature and dishes in the case of food—are constructed for each respondent. These lists include twenty-eight items that represent the geographic area of residence at selected ages, items representing areas where the respondent never lived, and items presumed to transcend any particular region. Respondents are asked to indicate their degree of familiarity and preference for each author or dish.

Respondents who receive the music module are presented with a list of music genres and asked to indicate their familiarity and preference for each type; this list is identical for all respondents. Each respondent then assesses a smaller set of genres, in an effort to more precisely understand his or her knowledge of the variation within a given genre. In some cases (when classical, jazz, country or dance music are presented), respondents are offered the chance to hear sound clips that represent particular sub-genres. The views-on-the-world topical module consists of eighteen Likert-scale items that complement the community questions asked of all respondents. These items tap respondents' views of the world (e.g., complexity, optimism, and altruism) and include a subset of Internet-related items.<sup>iv</sup>

Though each respondent initially receives a single topical module, they are also offered the option to complete the remaining topical modules and to comment on two open-ended questions. The survey closes with a screen that informs a randomly selected subgroup of the sample that they have been selected to receive a gift for participating in the survey. As a further incentive, all respondents are offered a customized set of web links (URLs) that have been selected to reflect their individual responses to questionnaire items, including geographic areas they have lived in and their leisure time activities and interests.

### **III.B International Respondent Survey**

This survey instrument is designed for adult respondents who are neither US nor Canadian residents or citizens. In theory, it is easy to imagine a survey instrument for all respondents with a similar structure as that used for US and Canadian respondents. However, early in the project development process it was decided that the Survey2000 lacked the resources necessary to develop parallel instruments for all international respondents. Moreover, the organizational apparatus of the NGS, which is seen as a direct means to counter the bias inherent in a web survey, that is central to the US and Canadian adult survey, loses some of its efficacy outside North America.

Nonetheless, Survey2000 is a *worldwide* web survey. Defined in this manner, Survey2000 could hardly ignore respondents from outside North America or force them to respond to an instrument that showed little sensitivity to a significant subset of respondents. Thus, the decision was made to turn a potential limitation of Survey2000 into an advantage. Survey2000 offers a unique opportunity to empirically test of the claim that American cultural imperialism has cause the emergence of a global McCulture, at the expense of regional and national cultural diversity. Survey2000 over-sampled well-educated and well-off respondents and this bias is likely to be more acute among international respondents. Furthermore, the most well educated and well-off segments of the world's population are most likely to have been exposed to and adopted cultural hallmarks of North America. To the extent that North American culture has left a hegemonic footprint, it should therefore be particularly apparent among Survey2000 respondents.

To consider this question the international respondent instrument begins with a standard set of demographic questions, residence, citizenship, age and gender, as well as marital status, household composition, educational attainment, and employment status. The international respondents also receive the same set of Internet use and access questions as North American respondents. The mobility history for international respondents is limited to current residence and place of birth. In addition, respondents are asked if they have ever visited or lived in the US or any other country than their current country of residence. International respondents receive the same set of questions regarding their social world, including contact with friends and relatives, group membership, political participation, and the Likert-scale community items.

The interests and perspectives section of the international survey is necessarily quite different from the US and Canadian instrument, which customizes the literature and food modules according to each respondent's personal geographic mobility history.<sup>v</sup> The international survey does not randomly allocate respondents to a single topical module, but instead gives each respondent an abbreviated version of the literature, food, and music modules. Each international respondent is provided a list of eight North American authors and eight North American dishes drawn from a pool of items presumed to transcend North American regional culture. The respondents receive the music topical module in its entirety. These items clearly do not reflect the diversity of world culture, though this design is well suited to mapping the spread of North American culture among a certain segment of the world.

### **III.C Youth Survey**

All respondents under the age of sixteen, regardless of nationality, are directed to the youth survey. The survey content differs substantially depending on whether the respondent is between the ages of thirteen and fifteen or twelve and younger. All children are asked to request parental permission to complete the survey, and then queried as to gender, citizenship, current residence (including zip or postal code), where they are completing the survey (e.g., home, school, parents workplace) and languages regularly spoken at home. Youth respondents are also presented with an abbreviated version of the mobility history with questions focusing on length of residence in current location, total number of residences occupied and place of birth. The social world set of questions for youth respondents focuses on household composition and activities undertaken with parents and guardians. Children aged thirteen through fifteen are also asked about parental supervision, peer values, parental involvement in school activities and

neighborhood safety and solidarity. Youth respondents of all ages are given standard self-esteem items, which come from the child supplement to the Panel Study of Income Dynamics (PSID) for the younger children and from the National Longitudinal Study of Youth (NLSY79) for those age thirteen through fifteen. The older youth respondents' items also include measures of locus of control and propensities toward risk-taking. Finally, the older youth respondents also receive a short set of items designed to measure attitudes toward the future, with a special emphasis on the next millennium.

## **IV The Web-Based Survey Design**

As survey research has become increasingly sophisticated, social scientists, in particular sociologists, have become increasingly aware of the extent to which the findings of survey research are part and parcel of the social process and technology of data collection.<sup>vi</sup> Survey research requires social interaction, which in turn is sensitive to the technology upon which the process rests. Different survey research techniques face-to-face interviews, self-administered paper and pencil questionnaires, telephone interviews, CAPI, and CATI not only depend on different technologies, but also organize the social dynamics of data collection in a different fashion (Bratton and Newsted, 1995). Thus, a basic understanding of the technology behind a Web-based survey, such as Survey2000, is crucial to understanding the overall dynamics of this new means to collect social science data (Kehoe and Pitkow, 1996).<sup>vii</sup>

### **IV.A Program Features**

Survey2000 is delivered to the respondent population via a PERL script. The script takes each page as it is submitted, writes the data into an Oracle database, and determines which page should be presented next. Wherever possible, the script suppresses questions that have been implicitly answered or rendered irrelevant by earlier responses.

In the case of food and literature questions, using a web-based script supports flexibility that would never be possible in a self-administered paper survey. The lists of dishes and authors delivered to each respondent are custom-built based on the various locations the respondent reported living in during an earlier segment of the survey. PERL and Oracle were used for this project mostly as matters of convenience and familiarity. Because the data is in a conventional flat-file format, there is no application for Oracle's relational functionality.

One benefit of a customized survey approach is that it shortens survey time. For example, questions concerning the frequency with which an individual engages in specific Internet activities is not asked of those respondents who say that they are using the Internet for the first time to complete Survey2000. Similarly, respondents who say that they live alone are not queried as to whether specific relatives live in their household. The greatest efficiency gains are realized in the geographic mobility section.

### **IV.B Design Elements**

The design elements of Survey2000 are an essential feature of the project. Most fundamentally, the project's affiliation with the National Geographic Society is intended to

provide the project with a credibility and a sampling platform that few Web sites can offer. The NGS web site is well designed, regularly maintained, and attracts approximately 1.5 million “hits” per month. During the two-month period of data collection, a link to the survey was placed on the NGS home page. References to the survey site were also published in the NGS’s adult and children periodicals as well.

The NGS web design group used its considerable experience to enhance the aesthetics and functionality of the survey layout. Once a respondent begins the survey, the NGS logo remains on every page; however, there are no links to other pages or other sites to tempt the respondent into going elsewhere. The goal is not simply to capture respondents, but to engage and reward them for their participation. For example, the sidebar on the mobility history screens is customized for each respondent. When a respondent is asked where she lived at birth or at a given year the sidebar text reminds the respondent of the year in question and list three events that took place in that year.<sup>viii</sup> These facts are designed as prompts, but also as a way to make the process a bit more entertaining. Moreover, beyond the programming logic described above, an effort is made in the survey design to reduce respondent burden as well. For example, “check boxes” and “radio buttons” are used extensively to minimize the respondents’ keyboard input.

Pre-test results revealed that respondent burden might be too heavy if each respondent were expected to complete all four of the cultural modules. Thus, each respondent was given one of the four cultural modules; after completing the base survey and one of the four topical modules, each respondent received a thank you page, that included the option to continue the survey and respond to the three remaining topical modules. This approach might have had cost of its own with respect to possibilities for analysis across cultural domains; however, over half of the US adults (N=23,384) voluntarily continued after completing the base module. Thus, researchers can consider correlations across cultural domains – are individuals’ preferences in music and food driven by the same factors that affect their literary tastes?

## **V Sample Overview**

Tables 1 provides an overview of participation and completion of the entire Survey2000 sample. Over 80,000 surveys were initiated and just over 50,000 were completed. Adults living in or citizens of the US or Canada initiated the most surveys (N=45,951) and completed over 37,000 surveys. Adults from other part of the world comprised about one-fifth of the initiated surveys. Youth surveys were completed by 9,785 children in the US, 970 Canadian children, and 1,635 international youth. The overall survey completion rate was over 70% for all adults and almost 60% for children.

Just about half of the initiated surveys (40,612) were from US adult respondents. Combined with an 80.5% completion rate, this yields a sample of 32,688 complete US adult surveys. For many of the questions these data address, partial responses are of interest. Thus, a sample size of more than 32,688 respondents exists for relevant demographic and social capital measures that were the start of the survey instrument.

**Table 1. Survey2000 Participation and Completion by Survey Form**

---

---

	Number of Surveys Initiated	Number of Surveys Completed <sup>3)</sup>	Percent Completed
All Adults (age 16 or older) <sup>1)</sup>	65,676	47,176	71.8
All Youth (age 5 though 15)	14,339	7,761	54.1
US adults (age 16 or older)	40,612	32,688	80.5
US youth (age 5 though 15)	9,785	6,246	63.8
Canadian adults (age 16 or older)	5,339	4,403	82.5
Canadian youth (age 5 though 15)	970	633	65.3
Other international adults (age 16 or older)	13,613 <sup>2)</sup>	10,085	74.1
Other international youth (age 5 though 15)	1,635	882	53.9
Total	80,015	54,937	100.0

<sup>1)</sup> Includes 6,112 adult and 1,949 youth respondents who dropped out prior to reporting residence.

<sup>2)</sup> Total includes 961 (7.1% of total) respondents who are US or Canadian citizens living abroad.

<sup>3)</sup> A survey is considered completed if the respondent completed the base module to the point at which the first of four cultural modules began. See Table 1B for a more detailed breakdown of completion rates for US adults.

## V.A The Demographics of Participation

As noted above, for many applications, the critical issue concerning Survey2000 is the extent to which one can make generalizations from this sample to larger populations. The results presented in Table 2, which compares the Survey2000 sample with recent GSS surveys and US Census Bureau statistics concerning central demographic variables, ought to be viewed in two ways. First, these results indicate the extent to which the Survey2000 results will need to be statistically adjusted to correspond to the demographics of the US population and the Internet population within the US. Second, the Survey2000 sample provides some insight into the magnitude of the difference between the general US population and its Internet population.

**Table 2 Demographics of the Survey2000 Adults (Age > 18) sample compared to the 1996 and 1993 General Social Surveys and Census Bureau Statistics <sup>1)</sup>**

	Survey2000 <sup>2)</sup>		1996 General Social Survey		1993 General Social Survey		1997/1998 Census Bureau <sup>3) 4)</sup>	
	%	N	%	N	%	N	%	N (in thousands)
<b>Gender</b>								
Female	48.9	15,147	55.7	1,614	57.3	918	51.9	100,954
Male	51.1	15,801	44.3	1,283	42.7	683	48.1	93,474
<b>Median age</b>	38 years		44 years		43 years		40-44 years	
<b>Race</b>								
Black	1.4	428	13.9	402	11.2	179	11.6	22,590
White	94.5	29,004	80.9	2,344	83.9	1,343	84.0	163,368
Other	4.1	1,268	5.2	151	4.9	79	4.4	8,472
<b>Education</b>								
Less than HS degree	0.9	292	15.2	441	18.1	289	17.9	35,246
HS degree	31.9	9,882	54.1	1,567	52.5	840	52.9	104,334
Associate's degree	7.8	2,421	6.7	194	6.2	99	7.1	13,996
Bachelor's degree	34.1	10,569	16.3	471	15.8	253	15.2	30,087
Graduate degree	25.2	7,785	7.7	224	7.4	118	7.0	13,750

<sup>1)</sup> Sample is restricted to age 19 or older to facilitate GSS comparison for all data sources.

<sup>2)</sup> There were also 713 respondents who did not provide information on race.

<sup>3)</sup> Education numbers include those 18 years and older

<sup>4)</sup> Data Sources: Gender data for the year 1997 – U.S. Census Bureau, *The Official Statistics, Statistical Abstract of the United States: 1998*, September 30, 1998, No. 15, page 16. Race data for the year 1997 – U.S. Census Bureau, *The Official Statistics, Statistical Abstract of the United States: 1998*, September 30, 1998, No. 22, page 22. Education data for the year 1998 – U.S. Census Bureau, Population Division, *Current Population Report, Educational Attainment in the United States: Detailed Tables March 1998*, October, 1998, pages 2-6.

Recent findings indicate that the gender gap in Internet use for U.S. adults has essentially disappeared, and trends towards equal access by education, income, and race have also been noted (Media Metrix, WIRED, July 1999; Clemente 1998; Katz 1997). Particularly concerning gender, our results corroborate these findings. This comparison shows, for example, that while just over half (50.7%) of the Survey2000 sample is male, female respondents constitute the majority in the 1996 (55.7%) and 1993 (57.2%) GSS samples.<sup>x</sup> Further, the Survey2000 sample is considerably younger, with a median age of 38 years than that estimated by the GSS (44 years in 1996 and 43 years in 1993).

The Survey2000 sample supports the widely held view that minorities are under-represented on the Internet: 92.5% of the respondents are white.<sup>x</sup> Only 1.5% of the US adult

surveys are from African-Americans. In subsequent efforts to generalize from Survey2000 to the broader US population weights will be developed to make the necessary statistical adjustments. But it should also be pointed out that the large sample size should make this possible. Though only 1.5% of the respondents are African-American, this amounts to 538 surveys. The 1993 GSS only has 179 African Americans, while the GSS African-American over samples in 1982 and 1987 include 354 and 353 African-American respondents respectively. Even in 1996, when the GSS sample was doubled, African-Americans only number 402. Thus, the actual proportions across categories are not as important as the number of respondents within each cell. With the large number of African-Americans who completed Survey2000 we will be able to draw inferences using statistical theory.

Finally, Table 2 indicates a large difference between the educational makeup of the Survey2000 sample and the US population at-large. Only nine-tenths of a percent of Survey2000 respondents have less than a high school degree, as compared to 15.2% of the 1996 GSS sample and 18.1% of the 1998 GSS sample. The GSS estimates are quite consistent with Census Bureau 1998 statistics that indicate that 17.9% of the US population age 18 or older has less than a high school degree. Also, the proportion of Survey2000 respondents with a high school degree but no post-secondary degree (31.9%) is considerably lower than that found in the 1996 GSS (54.1%) and the 1993 GSS (52.5%). Correspondingly, respondents with post-secondary degrees are over represented in the Survey2000 sample. The proportion of Survey2000 respondents with an Associate's degree is quite similar to the Census Bureau statistics; but the proportion with a Bachelor's degree (34.1%) is roughly double that provided by the Census Bureau. The proportion of Survey2000 respondents with a graduate degree (25.2%) is more than three times the official Census population estimates. Once again, this means that weighting will be required to make any generalizations from the sample to the US population at large, but in and of themselves these numbers reveal a great deal about the educational background of the Internet community. These data are instrumental in assisting researchers' early attempts to represent the population of Internet users.

For purposes of analysis as well as the construction of weights, the distribution of cases across a single demographic variable only tells part of the story. One also must consider the number of cases sharing the same combination of several characteristics, i.e., the cases occupying a single cell in a joint contingency table constructed using several variables. Table 3 compares the Survey2000 and 1996 GSS samples according to this criterion. Each of the 120 cells in this table represent a unique combination of gender, race, age cohort and educational attainment. The upper entry in each cell indicates the number of Survey2000 respondents with this particular combination of characteristics, while the lower entry states the number for the GSS. At least one Survey2000 respondent reports each combination of characteristics, however in 52 of the cells the number of cases drops below the value of thirty, a value often considered a minimum needed for the analysis of subgroups. Though a smaller number of cells with few observations would certainly be preferred, it should be noted that 86 of the cells in Table 3 contain fewer than thirty observations when the 1996 GSS is considered. Moreover, the 1996 1994 and samples are approximately double the size of previous GSS samples. For example, considering the 1993 GSS sample, which contains many of the cultural items used in Survey2000, 101 of the 120 cells contain fewer than thirty observations.

**Table 3: Absolute numbers of cases with particular combinations of demographic characteristics in the Survey2000 (upper cell entry) and 1996 GSS samples (lower cell entry)**

Survey2000 sample GSS sample	Male			Female		
	White	Black	Other	White	Black	Other
<b>Under 25</b>						
HS or less	1042	33	118	1074	32	101
	61	11	9	49	22	8
Some college	1033	22	87	1161	28	84
	36	3	3	45	12	5
BA degree	475	5	37	622	19	52
	10	-	-	9	1	1
Graduate degree	28	1	8	47	1	11
	-	-	-	-	-	-
<b>25-34</b>						
HS or less	257	5	13	238	10	14
	73	7	11	111	32	7
Some college	1247	22	50	1112	43	74
	95	16	6	88	22	10
BA degree	1983	22	110	2000	31	114
	43	4	5	66	15	5
Graduate degree	906	15	124	1027	28	85
	14	-	3	13	1	1
<b>35-44</b>						
HS or less	372	4	10	398	3	18
	107	19	4	111	25	12
Some college	1304	21	49	1467	35	63
	74	14	5	88	17	6
BA degree	1380	9	51	1428	24	41
	48	3	5	57	10	7
Graduate degree	1106	13	65	1011	20	41
	25	2	3	30	2	2
<b>45-54</b>						
HS or less	267	6	7	274	6	8
	68	9	4	81	33	3
Some college	1088	19	33	1158	21	33
	69	5	1	66	13	2
BA degree	1158	6	35	1087	15	38
	48	4	5	50	4	2
Graduate degree	1383	18	46	1179	16	30
	28	1	1	32	4	1
<b>55 and older</b>						
HS or less	330	2	1	243	1	3
	158	25	4	241	48	4
Some college	857	8	10	668	2	15
	56	3	-	81	4	-
BA degree	883	8	13	519	7	7

	30	1	2	34	1	1
Graduate degree	1155	10	25	652	7	10
	33	4	3	19	2	-

---

## V.B Survey Completion and Attrition

Although one can not speak of a traditional interviewer effect in a web-based survey, there is a technologically-mediated relationship between the respondent and the survey. Many factors shape this relationship including: the respondent's prior experience with the web, the quality of the internet service provide (ISP), the type of hardware and software at the site of the respondent as well as at the location of the server, the speed and design of the survey. All of this comes as no surprise

**Table 4a: Completion Rates for U.S. Adults by Sex and Age**

	Sex				Age <sup>1)</sup>			
	Men	Women	Did not answer	Total	38 or younger	Older than 38	Total	
Dropped out during base section (before first culture module)	19.9%	18.7%	28.8%	19.4%	19.0%	19.9%	19.4%	
Completed base section and one culture module	25.0%	20.9%	23.8%	23.0%	21.5%	24.6%	23.0%	
Completed base and two or three culture modules	7.4%	8.6%	5.2%	8.0%	8.4%	7.6%	8.0%	
Completed base and all four culture modules	47.7%	51.8%	42.2%	49.6%	51.1%	47.9%	49.6%	
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
Number of cases	20577	19598	445	40620	20646	19996	40642	
	$\chi^2$	160.85**				75.16**		

\*\*  $\chi^2$  is significant at  $p < .01$ , \*  $\chi^2$  is significant at  $p < .05$

<sup>1)</sup> There are no missing data for age for all 40,642 US adult respondents because birth year was required for respondents to route them to the appropriate instrument.

**Table 4b: Completion Rates for U.S. Adults by Race and Education**

	Race					Education					Tot	
	Black	White	Other	Did not answer	Total	HS degree or less	Some College	Bachelors degree	Graduate degree	Did not answer		
Dropped out during base section (before first culture module)	25.2%	16.5%	22.8%	24.7%	17.1%	22.3%	16.8%	15.3%	16.2%	35.8%	17	
Completed base section and one culture module	25.7%	23.5%	26.6%	23.5%	23.7%	26.7%	23.1%	22.5%	24.3%	23.6%	23	
Completed base and two or three culture modules	6.0%	8.3%	6.8%	6.2%	8.2%	8.1%	8.4%	8.7%	7.5%	6.7%	8	
Completed base and all four culture modules	43.1%	51.6%	43.7%	45.6%	51.0%	42.9%	51.7%	53.6%	51.9%	33.9%	51	
Total	100%	100.0%	100.0%	100%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100	
Number of cases	615	36320	1783	790	39508	5067	12157	12474	9300	508	39	
	$\chi^2$	144.08**					367.92**					

\*\*  $\chi^2$  is significant at  $p < .01$ , \*  $\chi^2$  is significant at  $p < .05$

**Table 5a: Completion Rates for U.S. Adults by Server and Platform**

	Server			Platform				Total
	Mountain	Magma	Total	Windows	Mac	UNIX	Other	
Dropped out during base section (before first culture module)	33.7%	15.8%	19.4%	18.5%	28.8%	11.9%	17.5%	19.4%
Completed base section and one culture module	21.0%	23.6%	23.0%	23.5%	18.8%	18.6%	23.6%	23.0%
Completed base and two or three culture modules	7.7%	8.0%	8.0%	5.9%	8.3%	6.2%	4.1%	8.0%
Completed base and all four culture modules	37.6%	52.6%	49.6%	49.7%	46.3%	65.4%	53.0%	49.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Number of cases	8318	32324	40642	35516	4046	269	811	40642
	$\chi^2$	1416.34**			300.96**			

\*\*  $\chi^2$  is significant at  $p < .01$ , \*  $\chi^2$  is significant at  $p < .05$

**Table 5b: Completion Rates for U.S. Adults by Browser**

	Browser					Other
	Netscape alone	MS Internet Explorer alone	MS Internet Explorer w/ AOL	MS Internet Explorer w/ MSN		
Dropped out during base section (before first culture module)	18.5%	18.5%	26.8%	13.7%		17.0%
Did base section and one culture Module	20.6%	25.4%	26.1%	22.7%		21.1%
Completed base and two or three culture modules	4.7%	9.9%	16.2%	4.7%		4.2%
Completed base and all four culture modules	56.2%	46.3%	30.9%	58.9%		57.7%
Total	100.0%	100.0%	100.0%	100.0%		100.0%
Number of cases	18521	13218	5764	1766		1373
	$\chi^2$	1810.97**				

\*\*  $\chi^2$  is significant at  $p < .01$ , \*  $\chi^2$  is significant at  $p < .05$

## VI Survey Logic and Skip Patterns

A distinguishing characteristic of Survey2000 is its complex logic and skip patterns. Computer assisted interviewing allows a survey instrument to be customized and tailored to the individual respondents. This feature is used in Survey2000 to expedite the collection of mobility history data and to create measures of cultural tastes in food and literature that reflect each respondent's personal mobility history. The complexity of the survey instrument leaves its mark on the data set as well. Thus, to successfully work with the data, users should have a good understanding of the logic and skip patterns built into the survey.

Just as with a paper survey, Survey2000 presents individual items on pages, whereby in a web survey a page indicates a screen containing text and one or more survey items. In the following discussion font formats are used to distinguish the "handles" used to identify from the **dataset variable names** corresponding to items associated with a given page.

[index.html](#) Users began with this page after being referred from "Begin Survey2000" link at [www.nationalgeographic.com](http://www.nationalgeographic.com). The page includes a brief description of the project and its goals, as well as links to a statement of confidentiality, credits and the youth version of Survey2000. A "Begin Survey 2000" button brings the respondent to:

**init**, which collects data on birth year, gender, continent of citizenship and country of citizenship (**byear, sex, citizen1, citizen2**). Data entered for birth year is subject to the following rules: If no **byear** is entered, the person is directed to return to this page. If a 2-digit year is entered, it's assumed to be short for "19xx". Birth years must be between 1890 and 1993. People who enter birth years from 1983-1993 are directed to the youth survey. Citizenship is queried at this point as well. If a person chooses 0 or more than one country of citizenship from the lists, the person is returned to this page to reselect a country. The lists are converted into continent (**citizen1**) and country (**citizen2**) values. Choices come from "[nations.html](#)," a separate file integrated into drop-down menus at run-time. When this page is submitted, a unique serial number is assigned which is used to track all other responses. Start time, browser, IP address, host domain name are logged as well. The latter two variables are intended for internal use only and are not distributed to protect the confidentiality of respondents.

### **demogspl**

This is a non-interactive splash screen which lets the respondent know that the series of pages collect demographic data.

### **demog1** (**lcontcu, lcouncu, lpostcu**)

This page lists six drop down menus corresponding to the continents/landmasses, North America, South America, Europe, Asia, Africa and Oceania. Each drop down menu includes a comprehensive list of countries and territories associated with the particular continent or landmass. Country choices come from "[locations.html](#)," a separate file integrated at run-time. Respondents are only allowed to select from a single menu. As with (**citizen1**) and (**citizen2**), the information on continent and country are stored in two variables (**lcontcu, lcouncu**) The public use data file contains two generated variables that combine

continent and country residence data into an alphanumeric variable (`cname`) and a numeric (`cnamen`) variable, each of which is sufficient to uniquely identify an individual country. Each respondent is also asked to enter their zip code. Respondents who indicate that they are US or Canadian residents or who previously indicated that they are US or Canadian citizens are assigned a `survform` value of 1 at this time. This value subsequently tracks a respondent to the US and Canadian and Adult version of the survey, while other adult respondents are given a `survform` value of 2 indicating that they are to receive the International Adult version. A `survform` value of 3 is used for participants in the Youth Survey.

### **demog2**(`marstat`,`nperinhh`,`nkidinhh` )

This page contains basic questions about marital status and household size. The value for `marstat` is initialized to -1; respondents are then asked to select from a dropdown menu. A blank field, reported in the SPSS data file as system missing, indicates that a respondent has dropped out of the survey prior to this page. The number of people in the household (`nperinhh`) and number of kids in household (`nkidinhh`) are meant to be entered as numbers. If anything but numbers are entered in these fields, a -1 goes into the database for that field. In this case a blank field may be the result of attrition or item nonresponse.

### **demog3**(`race`,`ethnich`,`ancestry`,`school`,`educ`,`empstat` )

This CGI generated page contains more demographic questions. People who claimed US residency earlier get US-specific race (`race`) and Hispanic ethnicity questions. (`ethnich`) Canadian residents get Canadian ethnicity/ancestry questions based on the 1996 Canadian Census and these responses are stored as the variable `ethnich`. Both `race` and `ethnich` are initialized to -1 once this page is loaded. Respondents who are neither US/Canadian citizens or residents receive neither version of this question and their fields for these variables remain blank. Everyone, regardless of citizenship, gets same question about enrollment status (`school`). US residents get one list of choices about level education completed (`educ`). All others get a differently worded set of choices for the same (`educ`) variable. People from all countries get the employment status question (`empstat`). Each of these variables (`school`,`educ`,`empstat`) are initialized to -1 when the page is loaded.

### **occup**(`occcur`,`occpast`,`workever` )

This is a CGI generated page about the respondents' occupation. This page takes one of two forms. 1) Individuals who respond to the employment status question ("`empstat`" on the **demog3** page) that they are employed full-time, are employed part-time or are employed but temporarily not working due to illness, vacation or a strike, are asked to enter their current occupation, "`occcur`". This is an open-ended question and respondents are asked to "be as specific as possible." 2) Individuals who respond that they are currently unemployed but looking for work, in the military, homemakers, retired or are not working and not looking for work are asked if they have ever held a job for at least one year, "`workever`". Individuals who respond that they have worked in the past are then asked about their most recent occupation "`occpast`". This, too, is an open-ended question and respondents are asked to "be as specific as possible."

## **inetspl**

A non-interactive splash screen letting the respondents know that a series of questions about internet access and use are to follow.

### **inet1** (interhow, interdur )

Respondents are first asked about where they are completing, for example at home or at work (interhow). Individuals are also queried as to when they first used the internet (interdur). Individuals who respond that this is their first time skip ahead to the mobility splash screen (mobspl), while those with other internet experience are asked about the frequency with which they engage in various online activities.

### **intera1** (intera01, intera02, intera03, intera04, intera05 )

Questions on this page concern the frequency with which respondents engage in five different online activities (send/receive e-mail?, take part in mailing lists, such as listservs and other group e-mails? access digital libraries, newspapers, or magazines? take online college courses or pursue other educational opportunities? purchase products or services?) and usage.

### **intera2** (intera06, intera07, intera08, intera09, intera10 )

Questions on this page address five other online activities (surf Web sites for recreational purposes? participate in Usenet newsgroups? engage in chats, such as Internet Relay Chat? visit MUDs, MOOs, MUSHs or other text-based or graphic multiuser environments? play multiuser online games, such as Doom, Quake, sports, or cards?).

### **mobspl**

A non-interactive splash screen informing respondents of the upcoming questions about their mobility history.

### **mob1** (resoutus, usvisit, uslive, elselive, resheret, resarea )

This page is CGI generated (varying depending on whether or not the respondent is a US resident). Those who previously claimed US residency are asked if they ever lived outside the US (resoutus). Those from other countries are asked if they have ever visited or lived in the US or elsewhere (usvisit, uslive, elselive). Everyone, regardless of residence, is asked the residency questions “resheret” and “resarea”. Resheret asks the number of years an individual has lived at his or her current address. Resarea, then asks if the respondent has always lived within 30 miles (50 kilometers) of that address. If anyone fails to answer resarea, they are directed back to the page to answer it, as this information is subsequently used to construct subsequent mobility pages.

### **mob2** (elsecont, elsecoun, resaptn, resaread, resaptn2, hhmemsp, hhmemch1, hhmemch2, hhmemp, hhmempl, hhmemsib, hhmemor, hhmemnr )

This is also a CGI generated page with additional mobility questions. US residents who answered that they previously had lived outside the US (resoutus) or non-residents who had lived somewhere other than their current residence (elselive), are asked to clarify which continent (elsecont) they had lived on and which country (elsecoun) they had lived in. The list of continents/countries is from the page "[locations.html](#)". People who have always lived within a 30 mile area (resarea = 1) are asked how many different residences they have

lived in within this area (`resaptn`). People who have not always lived within a 30 mile area are asked the questions `resaread`”, how long they have lived in this area, and “`resaptn2`”, how many different residences they have occupied in this area. Finally, respondents who said that there are people living in their household in addition to themselves (`nperinhh` on page [demog2](#)), are asked the questions (`hnmemsp`, `hnmemch1`, `hnmemch2`, `hnmemp`, `hnmempl`, `hnmemsib`, `hnmemor`, `hnmemn`) to determine the relationship to other members of the household.

### **mob3**

(`hnmemsp`, `hnmemesp`, `hnmemch1`, `hnmemch2`, `hnmemp`, `hnmempl`, `hnmemsib`, `hnmemor`). These questions query which relatives live in the respondent’s immediate area (within 30 miles/50 kilometers).

### **loc00con** (`lcont00`, `lcoun00` )

This page is also CGI generated page and asks respondents where they were born. As with residence and citizenship, the respondent selects country or territory of birth from one of six drop-down menus corresponding to continents/land masses. Information regarding the continent/landmass is stored in `lcont00`, while the country territory information is recorded in the variable `lcoun00`. Respondents who were born in the US or Canada, receive followup up questions concerning their state/province (`lstate00`) and city (`lcity00`) of birth.

After the questions about birthplace several rules govern the remainder of the mobility questions.

- 1) If the respondent is a non-US and non-Canadian citizen they skip to [socspl](#).
- 2) If they answer yes to `resarea` (they have always lived within 30 miles of their current residence, they skip ahead to [socspl](#).
- 3) If the next year that would asked about is in the future they skip to [socspl](#).
- 4) If the year just asked about is the same year they moved to immediate area (`resareat`) or this address (`resheret`), they skip ahead to [socspl](#).

If none of these challenges are met, a similar page is presented for the following ages:

7 (`lcont07`, `lcoun07`), 14 (`lcont14`, `lcoun14`), 21 (`lcont21`, `lcoun21`), 28 (`lcont28`, `lcoun28`), 35 (`lcont35`, `lcoun35`), 42 (`lcont42`, `lcoun42`), 50 (`lcont50`, `lcoun50`), 60 (`lcont60`, `lcoun60`), 70 (`lcont70`, `lcoun70`), 80 (`lcont80`, `lcoun80`), 90 (`lcont90`, `lcoun90`) as appropriate.

Rules 3) and 4) above are employed in considering residence at each age. If it is determined that a respondent has not attained a certain age or that value `resarea` or `resheret` indicates that a respondent’s residence at that age is the same as their current residence, the survey skips forward to [socspl](#).

### **socspl**

This is a non-interactive splash screen informing respondents that the next series of questions concerns their social world.

**soccrels** ([percrels](#), [telcrels](#), [letcrels](#), [webcrels](#) )

This is a CGI generated page that queries respondents regarding the frequency of social contact with nearby relatives. Separate items consider the frequency of four modes of contact: personal visits, phone calls or faxes, letters and cards, and e-mail or other Internet contacts. Respondents who indicated that they had no relatives living within 30 miles/50 kilometers (i.e., those who did not check any of the boxes associated with [hnmemsp](#), [hnmemesp](#), [hnmemch1](#), [hnmemch2](#), [hnmemp](#), [hnmempl](#), [hnmemsib](#), or [hnmemor](#) on the mobility page [mob3](#)) are not presented this page.

**soccfres** ([percfres](#), [telcfres](#), [letcfres](#), [webcfres](#) )

A CGI generated page questioning the frequency of social contact with nearby friends. These questions are asked of all adult respondents and the same four modes of contact are queried as on page [soccrels](#).

**socfrels** ([perfrels](#), [telfrels](#), [letfrels](#), [webfrels](#) )

A CGI generated page questioning the frequency of social contact with faraway (living beyond 30 miles/50 kilometers) relatives. These questions are asked of all adult respondents and the same four modes of contact are queried as on page [soccrels](#).

**socffres** ([perffres](#), [telffres](#), [letffres](#), [webffres](#) )

A CGI generated page questioning the frequency of social contact with faraway (living beyond 30 miles/50 kilometers) friends. These questions are asked of all adult respondents and the same four modes of contact are queried as on page [soccrels](#).

**help** ([percrelh](#), [percfreh](#), [perfrelh](#), [perffreh](#) )

This page asks how often the respondent gives or receives help or assistance with near and faraway relatives and friends. These questions are also asked of all adult respondents.

**org(1,2,3,4)** ([org01](#) .. [org20](#) )

Four pages of questions about organizational involvement. Each page lists five types of organizations (e.g., fraternal groups, sports groups, nationality or ethnic groups, or literary, art, discussion or study groups) and asks respondents to indicate the extent to which they participate (“Not at all,” “Am a member,” “Am an active member”) in groups of each type. These questions are presented to all adult respondents.

**pol(1,2,3)** ([pol01](#) .. [pol13](#) )

Three pages of questions about political participation and involvement. Each page begins with the statement: “Does your social world include political participation? In the past year, have you...” Thirteen different types of participation, ranging from signing a petition, and serving as an officer for a club or organization to engaging on political discussion on the internet. These questions are presented to all adult respondents.

**valatt1** ([valatt1](#), [valatt2](#), [valatt3](#) )

This page contains three items designed to capture individuals values and attitudes about community. The items are offered in radio-button format, forming a 7-point Likert scale ranging from “Strongly Agree” to “Strongly Disagree”. All adult respondents receive this page.

### **valint1** (valint1, valint2, valint3, valint4 )

This page contains four items measuring attitudes towards the internet. The items are also offered in radio-button format, as a 7-point Likert scale ranging from “Strongly Agree” to “Strongly Disagree,” and are presented to all adult respondents.

### **rec(1..5)** (rec01 .. rec30 )

These five pages ask about various recreational interests and activities. Each page lists six activities (e.g., going to the movies, visiting the library, hunting, dining out) and respondents are to “...indicate how often you have participated in the following activities during the past year.” Response include: “Never,” “Sometimes” and “Often.” These items, too, are given to all adult respondents.

### **intspl**

A non-interactive splash screen letting the participant know that the next group of questions is going to be about their interests & perspectives.

### **cultknow** (knowfood, knowmus, knowlit )

All respondents are queried as to how much they know about food, music and literature. In each instance respondents are to use a radio button select: “Quite a bit,” “A fair amount,” or “Not a great deal.”

### **cultint** (intfood, intmus, intlit )

All respondents are then asked: “A part from your knowledge of these areas, how interested are you in music, food, and literature.” In this case individuals select “Very interested,” “Somewhat interested,” or “Not at all interested.”

### **branch**

This is not really a page. It is an internal indicator that they are into the final section of the survey, which includes the food, literature, music and values modules. Every time "branch" is encountered, the survey log record is consulted to see which of the four "modules" or "branches" this respondent has seen. If the answer is none, than one is chosen randomly and presented. If the answer equals one, the respondent is asked whether he/she would like to end the survey or continue with further modules. If a respondent chooses not to continue, he or she is linked directly to the reflections page (**reflect1**). If the respondent elects to continue, he or she is assigned an additional module, each of which terminates at the **branch** indicator. Upon reaching the **branch** indicator a second time, it is assumed respondent wants to finish the entire survey. After each module, the next branch is chosen at random.

### **music(1..4)** (music01 .. music20 )

These four pages question the respondents feelings about twenty genres of music. Respondents are to select from radio buttons next to each genre corresponding to the options: “Like it very much,” “Like it,” “Have mixed feelings,” “Dislike it,” “Dislike it very much,” and “Don't know much about it.” US and Canadian survey respondents continue on with the **musicfav** page. Respondents receiving the non-US and non-Canadian version of the survey only get these questions in the music branch.

### **musicfav** (`musicfav` )

This is a CGI generated page that unfortunately contains a programming error that dramatically reduces the value of the information contained in the `musicfav` variable. The intention was that respondents would be presented with a drop down menu generated from all genres marked "like very much" on **music(1..4)**. If this list of selected genres contains fewer than five items, all genres marked "like" are added to the list. Users were then to select their favorite genre from this list and were to receive detailed questions concerning the selected genre (see **musdet00** below). Due to the programming error, however, the items on the drop down menu were not selected according to this logic nor was the selection random. Rather, genres that happened to be on the first two pages of genres were far more likely to be placed on respondents' menu lists and those on the last two pages were far less likely to be placed on the menu lists. Despite the programming error, the information as to the relative preference for a particular genre (as recorded in `musicfav` is accurate, however it can not be said that this choice indicates a respondent's top preference among all genres. Also the programming error did not affect the transfer of the selected music type to the subgenre routines found in **musdet00** below.

### **musdet00**

(`mdeta,mdetain,mdetb,mdetbin,mdetc,mdetcin,mdetd,mdetdin,mdete,mdetein,mdetf,mdetfin,genre` )

This is a CGI generated page containing "detail" questions for 10 genres (Classical, Opera, Jazz, Country, Bluegrass, Hymns/Gospel, R&B, Rap, Dance/Electronica, and Heavy Metal). If the respondent's favorite type of music (`musicfav`) is one of these genres, then more information is requested about that `genre`. If it is not, one `genre` is randomly chosen from those hidden above, with preference given to genres "liked very much". If `genre` is one of four with RealAudio examples (Classical, Jazz, Country, Dance), respondent is sent to a page where they can listen to music pieces (see below, **outloud**). If not, a page of detail questions is generated and the respondent is asked "How do you feel about this subgenre?" and "Does it belong in category 'genre'?" The respondent's reaction to the first question is elicited through a drop down menu with the options: "Like it very much," "Like it," "Have mixed feelings," "Dislike it," and "Dislike it very much." The response to the first genre is stored in `mdeta`, to the second in `mdetb` and so on. Next to each subgenre is a series of three radio buttons, where respondents are to indicate: "Do you think this fits in the category..." followed by the genre stored in `genre`. Respondents are to indicate, "Yes," "No," or "Not Sure." These responses are then stored in the variables `mdetain` to `mdetfin`.

### **outloud** (`outloud` )

This page asks if respondent wants to use RealAudio. If so, they are sent to the RealAudio page, **mdetraXX**. If they do not want to use RealAudio, they are sent to **musdetXX** page where XX goes directly to detail questions about the genre.

### **mdetra(01,04,07,12)**

(`modeta,modetain,modetb,modetbin,modetc,modetcin,modetd,modetdin,modete,modetein,modetf,modetfin` )

The respondent is given 5 or 6 RealAudio clips within the given genre to evaluate on same scales as silent questions. After this page is submitted, respondent is sent to *musdetXX* to answer silent questions about same genre. The logic for storing responses is similar to that used in the silent version: preferences regarding the subgenres are stored in *modeta* to *modetf* , and response as to whether the subgenre fits the category are recorded in *modeta*in to *modetfin* .

*lit(0..6)*(*au01num*...*au28num*,*au01val*...*au28val* )

These seven CGI-generated pages elicit respondents' reactions to 28 authors. Most of these authors are associated with an individual state or province, though a particular subset of authors were selected as "universal authors" with a widespread reading audience in the US and Canada. For US and Canadian survey respondents these 28 authors are selected from the author lookup table, which includes a total of 252 authors. The authors assigned to a particular respondent are selected according to the following rules:

- 1) up to four authors are randomly selected from the state lived in at birth
- 2) up to four authors are randomly selected from the state lived in at age 14
- 3) up to four authors are randomly selected from the state lived in at age 21
- 4) up to four authors are randomly selected from the state currently lived in
- 5) up to four authors are randomly selected from a state chosen at random from other states lived in
- 6) four authors are randomly selected from the universal author list

(Note: fewer than four authors could be identified for some states. Even if any of the bases for selection are duplicated, no more than four authors are chosen from that state at this point.)

This list of authors is then padded by selecting random authors from all available authors until the total is 28. These choices are written in the form *XXYY* to the variables *au01num* .. *au28num* where *XX* is the state/province code and *YY* is the numeric identifier of the specific author from that state.

Each respondent then cycles through the authors 4 at a time, until all 28 authors have been assessed. Each page is headed with the statement: "Please consider these authors—some famous, some not—and indicate the degree of familiarity you have with each." Respondents select from radio buttons next to each author's name corresponding to the following choices: "Have never heard of this author," "Have heard of; haven't read," "Have read something by this author," "I have read lots by this author." Answers are stored in *au01val*...*au28val* corresponding to the author's sequential position among the 28 authors across the 7 pages.

Respondents who are neither US or Canadian citizens or residents, are presented with only two pages, containing a total of 8 authors randomly selected from the universal author list.

*food(0..6)*(*di01num* .. *di28num*, *di01val* .. *di28val* )

The logic for the food section is quite similar to that used for literature. The section consists of 7 CGI-generated pages listing a total of 28 dishes. These dishes are drawn from a total list of 173

dishes. As was the case with literature a subset of these dishes are considered to be universal to the US and Canada and not specific to any particular geographic area. However, in the case of food section the remainder of the dishes are oriented toward regions (combinations of states or provinces), and not individual states or provinces as was the case in the literature section. For respondents completing the US and Canadian forms of the survey, the 28 dishes are selected by the following rules:

- 1) four are randomly selected from region associated with the state lived in at birth
- 2) four are randomly selected from region associated with the state lived in at age 14
- 3) four are randomly selected from region associated with the state lived in at age 21
- 4) four are randomly selected from region associated with the current state lived in
- 5) four are randomly selected from region associated with a randomly selected other state lived in
- 6) four are randomly selected from the list of universal dishes.

(Note: fewer than four dishes could be identified for some regions. Even if any of the bases for selection are duplicated, no more than four authors are chosen from that state at this point.)

This list of dishes is padded by selecting random dishes from all available until the total is 28. Similar to the logic used for literature, the identifiers associated with the selected dishes are written in the form XXYY to the variables `di01num` .. `di28num` where XX is the region number and YY is the numeric identifier of the specific dish from that region. For other than US & Canadian citizens, 8 dishes are selected randomly from the universal food list.

Each respondent cycles through the 7 pages, each of which lists 4 dishes. A series of radio buttons are next to each dish and respondents are asked to: 4 at a time until all dishes have been assessed. Answers are stored in `dizzval` corresponding to `dizznum`.

**`valatt(2..4)`**(`valatt04`...`valatt15` )  
**`valint(2..3)`**(`valint05`...`valatt10` )

These five pages contain further questions about general community and internet-specific attitudes. The page format is similar to **`valatt1`** and **`valint1`**. The items are offered in radio-button format, forming a 7-point Likert scale ranging from “Strongly Agree” to “Strongly Disagree”. All respondents on this branch get all questions, regardless of survey form.

**`reflect1`**(`reflect1`,`reflect2` )

This page asks questions requiring free-response answers. Answers have a maximum length of 2000 characters for each question. Responses are silently truncated upon submission, if they exceed this length.

**`end`**

Respondents who successfully completed the survey, either by choosing to end the survey after one cultural branch or by completing all the branches are eligible to receive a gift from National

Geographic Online. This page determines gift status. If fewer than 200 gifts have been given and more than 100 adults have finished survey since last gift, the respondent is offered a gift. If they are to receive a gift, contact information and gift (book or CD) preference is requested. On submission, they are sent to [gift](#) page.

Individuals not eligible for a gift (and gift recipients after they provide contact information), are provided with a set of selected URLs based on answers to earlier responses. These are intended as a small show of appreciation for respondents' participation. URLs are selected from a full set of URLs according to the following criteria:

music: one URL for each genre "liked very much"  
food: several URLs if "intfood" response was "very interested"  
literature: several URLs if "intlitt" response was "very interested"  
recXX: one or more URLs for each activity participated in "often" if URL available  
state/province: one URL for each state or province lived in

### [gift](#)

Contact data is written to the database and sent by email it to NG Online gift administrators. At this time the "sincelast" gift counter is reset to zero and the "awarded" counter is incremented by one. At this point gift recipients are returned to the [page](#) and provide with a URL list as described above.

---

<sup>i</sup> The Survey2000 world-wide web site can be visited at [www.nationalgeographic.com](http://www.nationalgeographic.com).

<sup>ii</sup> Respondents' Internet protocol (IP) addresses are also recorded as part of the Survey2000 process, as is the case in all Internet connections. This information can uniquely identify connected machines (though not respondents) if the machine has a static IP address. However, most respondents entered Survey2000 through public access Internet providers that assign the same IP address to a large number of respondents. This safeguards the identity of individual respondents, but at the same time permits the analysis of correlations among respondents associated with the same IP address. In any event, IP addresses will not be distributed as part of the public use data file.

<sup>iii</sup> These items include: 1) I feel close to other people in my community. 2) My daily activities do not create anything worthwhile for my community. 3) My community is a source of comfort. 4) I feel a sense of community with the people I've met on the Internet. 5) I have made new friends by meeting people on the Internet. 6) The Internet has brought my immediate family closer together. 7) The Internet has brought my extended family closer together.

<sup>iv</sup> 1) The world is too complex for me. 2) I don't feel I belong to anything I'd call a community. 3) People who do a favor expect nothing in return. 4) I have something valuable to give to the world. 5) The world is becoming a better place for everyone. 6) I cannot make sense of what's going on in the world. 7) Society has stopped making progress. 8) People do not care about other people's problems. 9) I find it easy to predict what will happen next in society. 10) Society isn't improving for people like me. 11) I believe that people are kind. 12) I have nothing important to contribute to society. 13) Talking with people on the Internet is as safe as communicating with people in other ways. 14) The Internet has allowed me to communicate with all kinds of interesting people I otherwise would never have interacted with. 15) The Internet isolates people from one another. 16) I feel I belong to an on-line community on the Internet. 17) Information on the Internet is as trustworthy as information from television and newspapers. 18) I can find people who share my exact interests more easily on the Internet than I can in my daily life off-line.

<sup>v</sup> Not only is the mobility history for international respondents not as detailed as that collected for North American respondents, but also collecting the information on world literature and cuisine requisite to implement this design worldwide is beyond the scope of the Survey2000 project.

<sup>vi</sup> The significance of this relationship is noted quite clearly by Alfred Schutz (1971[1953]) in his discussion of how

---

the social scientist's field of inquiry fundamentally differs from that of the natural scientist: "His observational field, the social world, is not essentially structureless. It has a particular meaning and relevance structure for the human beings living, thinking and acting therein."

<sup>vii</sup> Another interesting Internet survey project is hosted by the Graphics, Visualization, and Usability Center of the Georgia Institute of Technology, which has been conducting web-based surveys for 5 years. The GVU's 10<sup>th</sup> survey offered cash incentives to respondents, advertised corporate sponsorship, and collected over 5000 responses. Whereas the GVU is focussed on market penetration of Internet technologies and the rise of Internet usage, the Survey2000 is also focussed on how mobility shapes community values and cultural awareness. While GVU data allows generalization about some features of the Internet population over time, the Survey2000 allows both generalization about the internet population over time and comparison with larger populations.

<sup>viii</sup> Two examples of the sidebars: For 1972: J. Edgar Hoover, controversial director of the U.S. Federal Bureau of Investigation dies. Arab terrorists massacre Israeli athletes at the XX Olympiad in Munich. U.S. first-class postage: 8 cents. For 1968: Martin Luther King, Jr., and Robert F. Kennedy are assassinated two months apart. Film director Stanley Kubrik releases 2001: A Space Odyssey. U.S. first-class postage: 6 cents.

<sup>ix</sup> A long standing issue with the GSS and other probability samples has been the overrepresentation of females (Smith 1979).

<sup>x</sup> Another 713 respondents did not provide information concerning race. Presumably, many of these were not white. However, even if one assumes that all of those who failed to identify with one of the race categories are not white, still more than 90% of the sample is white.