INTERNET SURVEY CASE STUDIES

In this chapter, we present examples of Internet surveys that were fielded by various organizations. Some of these case studies have appeared in the literature and some have not. We present them here to illustrate the range of Internet survey possibilities.

These case studies include probability samples of general populations, probability samples of closed populations,[1] and convenience samples. We included surveys that were constructed using a commercial survey software product,[2] surveys that were programmed from scratch, and surveys that were conducted by a commercial Web survey company.

Although each study represents only one specific implementation of a survey, as a group they all serve to demonstrate the range of Internet survey options that have been tried and continue to be in use.

---

[1]As noted previously in this report, with closed populations it is usually possible to contact survey respondents via e-mail; this is usually not true for general populations because "general" e-mail lists do not exist.

[2]We personally encountered a number of challenges, of a purely technical nature, with a survey software package. Although this experience was limited to a single software package, we believe that the same or similar issues could arise with other such products. The following are a few of the difficulties we encountered: Resizing the browser while the survey is in progress caused the survey to disappear from the screen; after clicking Reload/Refresh, the survey reappeared. It was difficult to make the software display specific error messages in a human-readable format. Displaying explanatory text without an accompanying question was not possible and required a time-consuming work-around. Each response option was preceded by a bullet that could not be removed (this can be seen in Figure 6.3—the bullet is superfluous because each response option also has a check box).

A SURVEY USING A PROBABILITY SAMPLE OF A CLOSED POPULATION

The USAF Surveys Branch, an office of the Air Force Personnel Center at Randolph Air Force Base, San Antonio, Texas, surveys Air Force active-duty personnel and their family members, Air Force civilian employees, Air National Guard and Air Force Reserve personnel and their dependents, and Air Force retirees. The USAF Surveys Branch previously conducted paper-based surveys via postal mail, but it has recently converted to conducting surveys entirely over the Internet. The office is staffed with only four people who are responsible for drafting, fielding, and analyzing surveys of more than 350,000 active-duty Air Force personnel located throughout the world.

The USAF Surveys Branch uses e-mail as the contact mode and the Web as the response mode. E-mail addresses are constructed using a standard Air Force e-mail address format, such as the following: (firstname.lastname@airforcebase.mil. About 70 percent of e-mails constructed in this way actually work, meaning that about 30 percent of the e-mails are returned as undeliverable (for example, the recipient may have chosen his nickname in place of his full first name in the e-mail address or may have added a middle initial). The Surveys Branch also uses e-mail for nonresponse follow-up.

The USAF surveys are administered via the Web. The Surveys Branch developed its own software in SAS (a statistical software package) that automates both instrument development and the generation of HTML code so that the survey instrument can be created easily and then posted on a Web server. Survey responses are subsequently automatically saved in an electronic format that makes it easy to import the data back into SAS for analysis.

Survey topics include personnel, workforce, and quality-of-life issues. The surveys are conducted at the direction of various Air Force organizations and commanders. Survey response rates generally are around 35 to 40 percent—roughly equivalent to the rates the organization achieved with paper-based surveys.

The success of this all-electronic approach is attributable to three factors. First, the standardized e-mail address system provides an easy means to contact a random sample from a closed population

that can be completely enumerated. Second, the USAF Surveys Branch has detailed information on its entire population of interest. Third, most of the population has ready access to computers that are fairly standardized, so respondents can reasonably be expected to have access to a Web-access instrument and browser and, therefore, other software problems are minimized.[3]

Under these conditions, all-electronic surveys have the potential to achieve their promise of faster results at lower cost. For example, in response to a congressional inquiry, the Surveys Branch completed an Air Force–wide survey (from initial instrument design through final analysis and reporting) in just 11 days, with its staff of just four full-time personnel. Without the Internet-based survey capabilities, the number of personnel needed to accomplish this task undoubtedly would have needed to be much higher.

AN INEXPENSIVE WEB SURVEY WITH A CONVENIENCE SAMPLE

In 2001, RAND fielded a survey to collect information about victims of sexual assault. The target population consisted of 18- to 35-year-old females who had experienced sexual assault in the past five years. A convenience sample of about 100 respondents was recruited through advertisements in college newspapers and notices posted on support-group Web sites. Respondents were required to call in to receive a password in order to access the Web survey. During that call, respondents were screened for eligibility and, if found eligible, given a password. To protect privacy, minimal information was collected during the initial call and that information and the information collected from the survey were not linked.

_____

[3]Interestingly, Charlie Hamilton, the head of the Surveys Branch, told us that after having conducted a number of surveys in this format, the USAF found that two specific occupational groups are difficult to reach electronically: military police and nurses. Neither of these groups, due to their occupational constraints, has easy access to computers with Internet hookups. The USAF Surveys Branch is currently attempting to overcome this problem by oversampling and using additional nonresponse follow-up, but it is not clear whether these efforts can overcome any resulting nonresponse bias.

The survey consisted of approximately 35 Likert-scale[4] questions and participants automatically received an Amazon.com $15 gift certificate upon completing the survey.

We include this case study as a good example of what can and cannot be easily done when fielding a Web survey that is constructed using only HTML and no commercial software. The study topic was clearly a very sensitive one and the target population is one that is generally difficult to identify and reach. The use of the Web and mass advertising was an efficient way to reach a relatively rare scarce population while allowing anonymous survey participation. Programming the ten-minute survey took about three days of a senior programmer's time and about five days of a mid-level programmer's time. Figures 6.1 and 6.2 show sample screens from this survey.

We'd like to start by asking you some basic background information:

Q.1 Please indicate your age:
(please type into the space below)

Q.2 What is your marital status?
(please select one)

○ Married
○ Not married but living with a long term partner
○ Never married
○ Separated
○ Divorced
○ Widowed

Figure 6.1—Basic Survey Questions Created Without Commercial Survey Software

_____

[4]A *Likert scale* enables respondents to express their level of agreement or disagreement with a particular statement by specifying a value from one to five.

People sometimes look to others for companionship, assistance, or other types of support. Please indicate how often each of the following kinds of support would be available to you if you needed it.

(please select one answer for each question)

How often would you have...	None of the Time	A Little of the Time	Some of the Time	Most of the Time	All of the Time
1. Someone to help you if you were confined to bed (had to stay in bed).	⦿ 1	○ 2	○ 3	○ 4	○ 5
2. Someone to take you to the doctor if you needed it.	○ 1	○ 2	⦿ 3	○ 4	○ 5
3. Someone who shows you love and affection.	⦿ 1	○ 2	○ 3	○ 4	○ 5
4. Someone who hugs you.	○ 1	○ 2	⦿ 3	○ 4	○ 5
5. Someone to get together with for relaxation.	⦿ 1	○ 2	○ 3	○ 4	○ 5

Figure 6.2—Matrix Questions Created Without Commercial Survey Software

The survey is password-protected, and because it was coded using only HTML, the survey instrument is relatively simple. No images are used in the survey, the transmission of data is not encrypted, and error validation is limited (for example, the input of a negative value for "age" will not result in an error or warning message).

If the respondent wished to complete the survey at another time after starting it, her earlier answers were lost. The survey could have been programmed to allow for respondents to temporarily save responses for completion at a later date. In this instance, however, the brevity of the survey did not seem to warrant the additional programming costs this option would have incurred.

The survey featured a button interface that led the respondent to the next screen, but none that allowed the respondent to go back to the previous screen. Upon omission of an item, an error message appeared on a separate screen that read, "Please supply an answer for question *xx*." To return to that question, the respondent needed to click the browser's Back button and then click the browser's Reload/Refresh button to redisplay the previous screen. The respondents were required to submit an answer to every question in

order to proceed to the next screen (otherwise a respondent could obtain the incentive gift certificate without answering a single question).

Despite the limitations, this is an example of a simple research survey that was fielded via the Web relatively inexpensively. The researchers required only a convenience sample, and using the Web provided an inexpensive vehicle for anonymous response. Using conventional methods, the researchers would have still recruited participants in the same manner. However, the participants either would have had to be interviewed over the telephone, a much more expensive proposition, or they would have had to provide significantly more identifying information (such as a name and address) so that a paper survey could be mailed to them, which may have deterred participation. Furthermore, use of the Web gave respondents who completed the survey the opportunity to automatically receive an electronic gift certificate, a type of incentive that preserved their anonymity.

Because there is no efficient way to reach this target population, contracting with a Web survey organization, such as Knowledge Networks or Harris Interactive, that can target small cross-sections of the population, would have been a cost-competitive alternative.

A SURVEY WITH A PROBABILITY SAMPLE OF A GENERAL POPULATION AND MULTIPLE RESPONSE MODES

RAND was commissioned in 2000 to study alternative policy options to help improve the military services' ability to recruit enlisted personnel from the college market in addition to the traditional high school market. In response to this commission, Asch (2001) designed a survey that inquired into respondents' desire to enroll in the enlisted ranks of the military under various policy scenarios. The scenarios included various hypothetical incentive packages, such as a monthly stipend while attending college and various levels of starting salaries, and a possible restriction to certain college majors that were desirable from the military point of view and a possible restriction to certain occupations within the military.

Because college dropouts were an important part of the target population, and no obvious strategy exists to reach dropouts, the follow-

ing strategy was adopted: The sample frame consisted of high school students who graduated or were graduating in 1999 or 2001 and who had also indicated that they intended to attend college. A list from which to draw the sample was commercially available. The sample consisted of 12,500 students graduating in 1999 and 1,750 students graduating in 2001.

With the survey being fielded in early 2001, it was possible to reach current high school students (high school graduation year 2001), current college students (high school graduation year 1999), and college drop-outs (high school graduation year 1999). The disadvantage with this strategy was that all students graduating in 1999 had to be approached via their parents' addresses from two years earlier (or the addresses at which they were living in 1999), which was bound to lead to large nonresponse rates. (Some parents wrote to say that their son or daughter had moved overseas and could not participate or had already joined the military, or to request paper versions of the survey.)

Respondents were initially contacted by U.S. mail and asked to respond on the Web. Only in the second mail follow-up were the respondents also supplied with a paper questionnaire and the option to return the questionnaire by U.S. mail. More details about the follow-ups and the entire survey timeline are given in Table 6.1.

Table 6.1

Example Timeline for Web/Mail Survey

Description	Response Option	Days
Wave 1: Initial letter to 3,000 students	Web only	0
Wave 1: Letter to parents	Web only	7
Wave 2: Initial letter to 11,250 students	Web only	7
Wave 2: Letter to parents	Web only	14
Wave 1: Phone follow-up	Web only	14–28
Waves 1 and 2: First mail follow-up	Web only	21
Waves 1 and 2: Second mail follow-up with paper survey	Mail or Web	36
Waves 1 and 2: Third mail follow-up—reminder postcard	Mail or Web	43–46
Waves 1 and 2: Fourth mail follow-up— replacement survey (including $3 gift certificate to 5,700 students)	Mail or Web	67
End fielding	—	90

Figure 6.3 shows one of the screens from this Web survey. The Web option appeared particularly attractive because the target population consisted of college-bound or college-educated young adults between 17 and 21 years old who were thought to be Internet savvy. Because of the large sample size, this mode was also thought to be cost-effective.

During the time the survey was in the field, several hundred respondents used either e-mail or an 800-number telephone line to contact
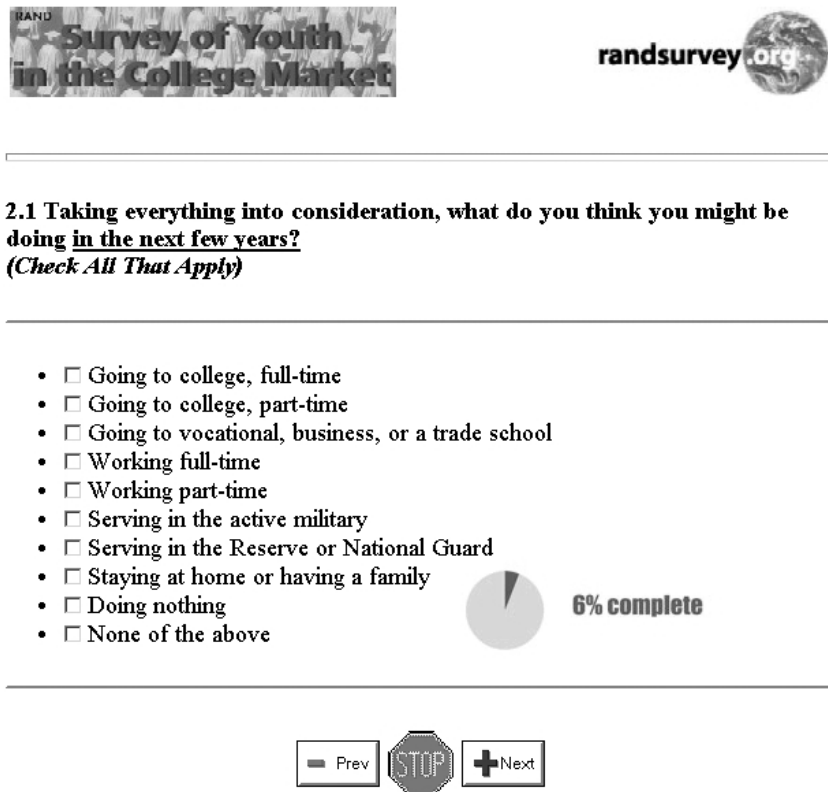

2.1 Taking everything into consideration, what do you think you might be doing in the next few years?
(Check All That Apply)

- □ Going to college, full-time
- □ Going to college, part-time
- □ Going to vocational, business, or a trade school
- □ Working full-time
- □ Working part-time
- □ Serving in the active military
- □ Serving in the Reserve or National Guard
- □ Staying at home or having a family
- □ Doing nothing
- □ None of the above

6% complete

Figure 6.3—Sample Screen from a Recruitment Survey

RAND. Several dozen respondents experienced password problems, however. Most of these problems were caused by the respondents mistaking the letter 'l' for the number '1' and vice versa, or mistaking the letter 'O' for the number '0' and vice versa. Some respondents experienced problems when using older AOL browsers.

The survey setup resulted in a few technical difficulties related to the first-time use of commercial software to construct the Web instrument. In a number of cases, specific features were desired that were not available in the software, necessitating cumbersome work-arounds. For example, the commercial software that was used provides generic error messages. RAND wanted to supply error messages that were customized to the individual question and that would warn the respondent that his or her answer was missing or did not meet a validation rule. The intent was to give the respondent an opportunity to change his or her answer, but not be forced into doing so.[5]

The survey produced 976 eligible responses via the Web and about 1,607 eligible responses by U.S. mail. For the Web survey, an additional 153 responses were found to be ineligible. Adjusting for the fraction of ineligibles, this yields a response rate of about 20.8 percent (62.2 percent of them by postal mail). This response rate should be seen in light of the fact that the majority of the respondents were no longer living at their parents' addresses.

It is noteworthy that there were more responses received by mail than via the Web, even though only the Web option was offered as a response mode in the initial survey mailing. Furthermore, because young adults are thought to be relatively Internet savvy and more likely to respond via the Web than by mail, it would suggest that when respondents are contacted via mail, providing them with a Web response mode is important. For this particular survey, eliminating the mail response clearly would have had a very negative effect on the overall response rate.

_____

[5]If the respondent's answer did not fulfill a validation rule and the respondent did not change the answer, the answer was reset to be missing. However, in some situations, the distinction between refusal, missing, and "don't know" may be informative and therefore it is important to distinguish among these categories.

To field this survey on the Web, the following costs over and above the costs for a mail survey were incurred: Programming the Web instrument required eight days of programming work; putting the survey on a Web server and related tasks required another 1.5 days of work; and managing the added amount of interaction with respondents required an additional ten days of survey coordinator time. However, for each survey returned via the Web rather than through the mail, approximately $7 per survey was saved in editing and data entry work.[6] Web responses received in the last couple of days before a follow-up mailing also saved $10 per mailing (including package preparation and postage).

For this survey, the break-even point at which adding a Web survey option was cost neutral occurred at about 580 Web surveys for eligible respondents.[7] This Web survey had 976 eligible responses, and we estimate that roughly $2,000 was saved by using the Web response option.[8] These savings mostly represent the eliminated cost of editing and data entry; only a fraction of the savings is from the eliminated cost for mailings.

A SURVEY BY A COMMERCIAL WEB SURVEY FIRM USING A CONVENIENCE SAMPLE ADJUSTED FOR SELF-SELECTION

As noted earlier in this report, Harris Interactive is an organization specializing in Web-based surveys. Harris Interactive solicits participants via Web advertisements and by other means, and maintains a database of millions of potential survey participants. For any particular survey effort, a sample of Harris's panel is asked to participate. The sample constitutes a convenience sample. Figure 6.4 shows a password screen and Figure 6.5 shows a survey question screen for a Harris Interactive survey.

---

[6]Open-ended questions were not coded on the paper surveys because of cost reasons; if they had been, the savings would have been even higher.

[7]The break-even number heavily depends on the programming time. All else held constant, if the programming work takes five days, the break-even point is at 510 Web surveys; for 15 days of programming work, the break-even point is at 730 Web surveys.

[8]This estimate is associated with considerable uncertainty and should not be taken too literally. This number does not include any fees or revenues on the part of RAND.

Screen shot courtesy of Harris Interactive.

Figure 6.4—Password Screen in a Harris Interactive Survey

Here, we describe a particular survey that Harris Interactive conducted regarding Californians' attitudes toward health care and health care providers, which was funded by the California Health Care Foundation (Schonlau et al., 2001).

Harris Interactive selected 70,932 e-mail addresses of California residents from its database of people who had volunteered to receive surveys and who had not recently received a Harris survey. A total of 81.5 percent of the selected e-mail addresses were chosen at random from all California residents in the database. The remainder was selected from specific California subpopulations in which oversampling was desired, including subpopulations of various racial and ethnic minorities, those aged 65 years and over, and respondents with low annual incomes (less than $15,000 a year).

The e-mail did not contain the survey but pointed to a password-protected Web page containing the survey. The Web page was accessible only to people with an individual password supplied in an e-mail. The survey was sent in two waves in early 2000 and potential

Screen shot courtesy of Harris Interactive.

Figure 6.5—Matrix Question in a Harris Interactive Survey

respondents had almost a month to respond after which the site was shut down. Nonrespondents received one e-mail reminder.

Of the 70,932 persons to whom an e-mail was sent, 2 percent started the survey and did not finish it and 12 percent completed the survey. Only 234 respondents were not eligible either because they were not 18 years or older or they did not reside in California, resulting in 8,195 eligible completes.

Because this type of Web survey does not employ a probability sample, weights are derived exclusively through post-stratification. The stratification matched the Current Population Survey (CPS) for California within race for gender, age, income, and health insurance, and, in addition, for variables derived from the propensity score.

Propensity scoring is a statistical technique (Rosenbaum and Rubin, 1983 and 1984) for comparing two populations. In essence, propensity scoring attempts to make two populations comparable by simultaneously controlling for all variables that are thought to affect the comparison. The Harris Interactive questionnaire included questions that measure general attitudes thought to differ between the on-line

and the general population. These attitudinal questions are then used for propensity scoring.

Initially, a Web survey and an RDD reference survey containing the same attitudinal questions are conducted. Propensity scores are obtained by performing a logistic regression on the variables representing attitudinal questions, using an indicator variable (the Web survey or RDD reference survey) as the outcome variable. Respondents for both surveys are sorted into five bins according to the propensity scores. Propensity weights are assigned such that the Web survey's weighted proportion of respondents in each bin matches the reference survey's proportion in each bin. Harris Interactive pioneered using propensity scores as weights (Terhanian et al., 2001). Note that propensity scores can be assigned to any subsequent Web survey that contains the attitudinal questions.

The weighting attempts to adjust for demographic differences of Web respondents compared with the general California population, and also for attitudinal differences of Web users compared with the general population. The key is that, at least in theory, the attitudinal variables compensate for the selection bias resulting from a Web sample consisting of Internet respondents only. The success of this method hinges on several assumptions: (1) the attitudinal variables capture the differences between the Internet population and the general population adequately; (2) no substantial bias is introduced in the event the RDD reference survey (usually targeting the entire U.S. population) is not identical to the target population (which may be a subset of the U.S. population); and (3) no substantial bias is introduced by the fact that the reference RDD typically is conducted a few weeks earlier than the Web survey.

Both Harris Interactive and RAND conducted identical surveys. However, RAND conducted an RDD phone survey whereas Harris conducted a Web survey. We compared several demographic variables for both the RDD and the Web survey with the CPS for California. It turned out that males were overrepresented in the RDD survey (57.5 percent) compared with the CPS (48.9 percent) and Web survey (46.6 percent). Only 9.7 percent of the Harris Interactive sample was Hispanic, compared with 25.1 percent of the CPS and 27.5 percent of the RDD sample. The Harris Interactive survey was conducted in English only, whereas the RDD survey was conducted

in both English and Spanish. The Web survey respondents tended to be better educated than Californians overall, based on the CPS.

Weighted-response estimates for nondemographic questions for the RDD and Web survey were different for some questions and similar for others. It was not possible to predict which questions in the two surveys would elicit similar answers. It was clear, however, that questions related to the Internet would yield very different answers. For example, the Web survey estimated that 84 percent of the California population used the Internet daily as opposed to a 24-percent estimate with the RDD survey.

We considered two subpopulations: (1) white respondents who were less than 55 years old, with at least some college, and with an annual income greater than $40,000 and (2) people who use the Internet daily. The estimated attitudes of Californians about health care and health care providers, based on the Web survey and the RDD survey, were closer in agreement for these subpopulations than the estimated attitudes in the general population overall.

We do not consider the RDD estimate to be the "right" estimate. However, it is comforting to obtain the same estimate from two different survey modes. To that extent, we believe the challenge remains to systematically explore the circumstances under which these two modes yield the same estimate.

TWO SURVEYS USING PROBABILITY SAMPLES FROM CLOSED POPULATIONS

The following subsections discuss two case studies of surveys using closed populations.

Prodigy Survey

Werner et al. (1995) conducted a study with an interesting application of e-mail surveys to public opinion tracking. From early 1993

through early 1995, Prodigy[9] conducted 84 e-mail surveys of their customers who were given seven questions about politics and economic issues. Of particular interest, the respondents were asked to "rate the overall job Bill Clinton is doing as president," which Prodigy then compared with the replies to a similar question posed by the Gallup Poll.

Because this survey was conducted with the cooperation of Prodigy, a complete list of e-mail addresses of the subscriber population was available. The survey was sent to a stratified random sample (by age, geographic region, and gender) of Prodigy subscribers with the goal of obtaining 1,200 to 1,500 completed surveys. Unfortunately, Werner et al. do not provide any information about either the initial sample size or the resulting response rates.

The interesting outcome of this experiment is that the approval measure of presidential performance generated by the Prodigy survey and the Gallup Poll survey both tracked in a very similar fashion. This was true even though Prodigy subscribers differ from the U.S. population in that they are more likely to be male, Republican, and married; they are older than the U.S. population on average; and a higher percentage of Prodigy subscribers are college graduates. Although there was a constant difference between the trends in the two surveys, with the Prodigy poll results consistently five to ten percentage points lower than the Gallup Poll results, the trend-line patterns tracked together in a strikingly similar manner.

The Prodigy poll did attempt to adjust for some population differences by post-stratifying and developing weights in order to match the U.S. demographic distribution by age, gender, geographic region, and political party affiliation. The remaining bias can likely be attributed to differences between the Prodigy population and the general population.

The consistency of the trend pattern between the Prodigy and Gallup Poll results is striking. It is tantalizing to think that the additional differential between the Gallup Poll and an on-line poll could be ap-

_____

[9]Prodigy is a commercial provider of on-line electronic services, such as e-mail and Internet browsing. During the time period of this study, Prodigy had about 2 million subscribers of which about 1.6 million were of voting age.

propriately corrected for, perhaps using more-sophisticated methods (such as propensity scoring) to adjust for nonresponse or imputation methods to account for missing data.[10] Furthermore, these efforts could be greatly enhanced if a service such as Prodigy collected additional respondent demographic information as part of the service registration process.

Pew Research Center Survey

Flemming and Sonner (1999) describe a polling effort by the Pew Research Center (an independent opinion research group sponsored by the Pew Charitable Trust) that provides an interesting contrast to the Prodigy experiment. Pew recruited two sets of respondents. The first group was the "volunteer sample." It consisted of visitors to the center's Web page who took a voluntary poll, and who then provided their e-mail addresses and agreed to participate in future Web surveys. The second group was the "selected sample" who were initially contacted as part of a random selection in a nationally representative telephone survey. They agreed to participate in future on-line surveys and provided their e-mail addresses. These samples were then used in two on-line surveys conducted in November 1998 and April 1999. Concurrent with the on-line surveys, Pew also conducted national telephone surveys using identical questions but with different participants, which provided a unique opportunity to compare the results of the on-line polls with results from a conventional RDD telephone survey.

Pew surveyed on a wide variety of topics. As with the Prodigy survey, questions about national issues, election issues, and party and candidate preferences were posed, but unlike the Prodigy experiment, Pew results were cross-sectional only. In addition to the questions posed in the first Pew survey, the second Pew survey also asked about "social advances," "technical progress," and "the reasons for America's success" and respondents' opinions about the future of the country. In all of these categories, the researchers found some similarities between the on-line and telephone results, but more of-

_____

[10]In the Prodigy experiment, respondents for which either age or gender were unknown were simply dropped from the sample. This amounted to about 8 percent of the Prodigy population.

ten found differences between the two, even after reweighting to adjust for known demographic characteristics. Most important, the differences were not consistent in any identifiable dimension. Sometimes the on-line results reflected more-conservative thinking and other times more-liberal thinking, and sometimes they reflected a more-pessimistic outlook and other times a more-optimistic one.

Summary of the Two Surveys

Traditionally, random samples from the *entire* target population have been used to ensure that a sample is representative of the target population. The Prodigy sample was randomly sampled from Prodigy subscribers, but if Prodigy subscribers are different from the general population, estimates may be biased. In other words, the sample constitutes a random sample for the population of Prodigy subscribers, but only a convenience sample for the U.S. voting population.

The success of the Prodigy experiment has implications for the potential success of convenience samples in general. However, whereas the Prodigy experiment seems to offer the promise of being able to adjust for the bias of an on-line poll, the Pew results seem to indicate that the magnitude and direction of such a bias, for any given question, may not *a priori* be identifiable. These questions are also directly relevant to the Harris Interactive approach discussed earlier.