

# **The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples**

**Robert P. Berrens and Alok K. Bohara**

*Department of Economics, University of New Mexico,  
Albuquerque, NM 87131*

**Hank Jenkins-Smith**

*George Bush School of Government and Public Service,  
Texas A&M University,  
College Station, TX 77843*

**Carol Silva**

*Department of Political Science, Texas A&M University,  
College Station, TX 77843*

**David L. Weimer**

*Department of Political Science and La Follette School of Public Affairs,  
University of Wisconsin-Madison,  
Madison, WI 53706  
e-mail: weimer@lafollette.wisc.edu*

The Internet offers a number of advantages as a survey mode: low marginal cost per completed response, capabilities for providing respondents with large quantities of information, speed, and elimination of interviewer bias. Those seeking these advantages confront the problem of representativeness both in terms of coverage of the population and capabilities for drawing random samples. Two major strategies have been pursued commercially to develop the Internet as a survey mode. One strategy, used by Harris Interactive, involves assembling a large panel of willing respondents who can be sampled. Another strategy, used by Knowledge Networks, involves using random digit dialing (RDD) telephone methods to recruit households to a panel of Web-TV enabled respondents. Do these panels adequately deal with the problem of representativeness to be useful in political science research? The authors address this question with results from parallel surveys on global climate change and the Kyoto Protocol administered by telephone to a national probability sample and by Internet to samples of the Harris Interactive and Knowledge Networks

---

*Authors' note:* We thank the National Science Foundation (NSF Grant Number 9818108) for financial support; Harris Interactive and Knowledge Networks for their contributions of survey samples; John Bremer, Hui Li, and Zachary Talarek provided valuable assistance at various stages; Charles Franklin, Ken Goldstein, Dana Mukamel, William Howell, Aidan Vining, and John Witte, participants in the Public Affairs Seminar and the Methodology Workshop at the University of Wisconsin–Madison, and the faculty of the Public Policy and Management Department, Shih Hsin University, provided helpful comments. The opinions expressed are solely those of the authors.

Copyright 2003 by the Society for Political Methodology

panels. Knowledge and opinion questions generally show statistically significant but substantively modest difference across the modes. With inclusion of standard demographic controls, typical relational models of interest to political scientists produce similar estimates of parameters across modes. It thus appears that, with appropriate weighting, samples from these panels are sufficiently representative of the U.S. population to be reasonable alternatives in many applications to samples gathered through RDD telephone surveys.

## 1 Introduction

Behavioral research in political science relies heavily on surveys. Since the early 1970s, the increasing costs of surveying in-person and technological improvements in surveying by telephone have made the telephone the primary mode of administration. With the cost and difficulty of administering telephone surveys increasing and the rapid spread of Internet use among the U.S. population, it is not surprising that firms and academics have begun to administer surveys using Internet technology. The Internet offers a number of advantages relative to telephone administration: dramatically lower marginal costs of producing completed surveys, superior capability for providing information (including through visual displays) and for asking complex questions, and the minimization of interviewer bias. The as yet incomplete penetration of Internet use among the U.S. population, prohibitions against message broadcasting (spamming), and the absence of a mechanism such as random digit dialing (RDD) to obtain probability samples, however, raise concerns about whether Internet samples can be sufficiently representative to support social science research. We report here on a study that compares the characteristics of samples from two prominent commercial Internet panels (developed by Harris Interactive and Knowledge Networks) and a national probability sample of respondents to a telephone survey on knowledge and attitudes toward global climate change and a related international treaty (Kyoto Protocol). Our analysis indicates that, despite some differences, the Internet and telephone samples would lead to similar answers for a number of substantive questions of interest to political scientists.

The potential uses of the Internet fall into three categories demanding different levels of sample representativeness. First, Internet surveys might be used to estimate population characteristics (means and proportions). As classically formulated, the reliable inference of population characteristics requires true probability samples, suggesting that as currently organized, Internet surveys are ill-suited to serve this function unless supplemented with data from non-Internet sources. Second, Internet surveys might be used to investigate relationships among variables. True probability samples may not be necessary to make valid inferences about relationships, especially when the most important variables of interest are based on “treatments” that are randomly applied to respondents. Indeed, witness the extensive use of convenience samples, such as students in psychology courses or experimental economics laboratory sessions to test hypotheses implied by social science theories. Third, Internet surveys might be used to investigate methodological issues in survey design that can be treated reasonably as independent of mode. The low marginal cost of completed surveys facilitates the comparison of such design issues as question order and format, which are unlikely to be highly sensitive to the characteristics of the sample. Thus, Internet surveys may prove useful in investigating methodological issues and as components of pretests for surveys administered by other modes. The study we report here demonstrates the third use and provides a rare opportunity for assessing how much progress has been made toward the second use.

It is important to keep in mind that several trends are likely to make telephone surveys relatively *less* attractive in the future. One trend is the general decline in response rates from both telephone and in-person surveys in recent decades (Steeh 1981; Atrostic et al. 1999; de Leeuw 1999; Steeh et al. 2000). Achieving acceptable response rates is likely to require greater investments in training and supervision or the use of costly financial incentives for respondents. Technological trends also suggest increasing problems for telephone surveys. The introduction of new area codes increases the ratio of possible numbers to working numbers. This reduces the probability of reaching working residential numbers (21% in 1988 vs. 13% in 1998) through RDD, and hence increases costs by increasing wasted calls (Piekarski 1999). Meanwhile, increases in cellular telephone subscriptions, from 3.5 million by the end of 1989 to 86 million by the end of 1999 (CTIA 2000), may eventually lead to more subscribers “cutting the cord” and no longer having fixed lines in geographic sampling frames. Another trend is increasing competition for respondents from telemarketing, push-polls, solicitations for charities, and telephone scams. Finally, university-based surveys face the added difficulty of getting through increasingly detailed informed consent statements before respondents disconnect.

Our objective is to provide insight into potential uses of Internet-based surveys in social science research. We begin by noting the increasing representativeness of the population of Internet users and the commercial responses to opportunities offered by Internet surveying. After describing the structure and purpose of the survey on global climate change, we make several comparisons among survey modes. First, we compare the socioeconomic characteristics of respondents. Second, as concern about sampling bias is based on possible differences in knowledge, attitudes, and behaviors not directly observable, we compare the samples in terms of knowledge about global climate change, degree of engagement in the survey as measured by the use and assessment of differing levels of information, and political attitudes. Third, as the focus of much research concerns relationships among variables, we investigate the relationship between political ideology and environmental attitudes, including support for the Kyoto Protocol. We conclude with observations about likely current and future uses of Internet surveys.

## 2 Increasing Opportunity for Internet Surveys

Errors in surveys can stem from a number of sources: coverage, sampling, nonresponse, and measurement (Couper 2000, p. 466). Internet coverage of U.S. households, although currently much less complete than the approximately 95% telephone coverage, is steadily increasing. The commercial potential of the Internet creates strong incentives to reduce sampling, nonresponse, and measurement errors. Current trends suggest that the Internet will become a more viable survey mode in the future.

Internet use in the United States has been growing rapidly, and is becoming more demographically representative. As recently as 1995, only about 10% of households were connected to the Internet (eMarketer 2000, p. 26). Estimates of the current fraction of households with Internet connections range from 26% to 44% (eMarketer 2000, pp. 25–26). A survey conducted in early 1999 found that among adults over 18 years of age, 34.0% had used the Internet at some time and 42.4% had access to the Internet at either work or home (U.S. Bureau of the Census 1999, p. 582). A national survey conducted between May–June and November–December 2000 found that the fraction of U.S. adult men with Internet access rose from 50% to 58%, and the fraction of U.S. adult women with Internet access rose from 45% to 54% (Rainie et al. 2001, p. 2).

The population of adult Internet users in the United States has different demographic characteristics than the general population. On average, it is younger, better educated, more male, in households with higher income, and disproportionately white and Asian. These differences appear to be diminishing rapidly. As recently as 1997, women made up only about 17% of Internet users, whereas they now make up 49% of users (eMarketer 2000, p. 56). By the end of 2000, the fraction of age cohorts with Internet access were as follows: 18–29 years, 75%; 30–49 years, 65%; 50–64 years, 51%; and 65 years and older, 15% (Rainie et al. 2001, p. 2). As those older than 55 compose the fastest growing Internet age group, their currently substantial underrepresentation is likely to diminish well before it is inevitably reduced by the aging of current users in the 35–54 age group (eMarketer 2000, p. 49). Despite recent declines in the median household income of Internet users, those with household income below \$20,000 remain underrepresented, at 6% versus 19% in the population; those with household incomes above \$150,000 remain overrepresented, at 8% versus 4% (Mediamark Research, June 2000, as reported in eMarketer 2000, pp. 72–73). Large-sample mail surveys of U.S. households conducted in January 1999 and 2000 show substantial convergence in Internet access across ethnic groups, with African-Americans, whites, and Hispanics each gaining about 10 percentage points (Walsh et al. 2000, p. 2). Access estimates at the end of 2000 were: whites, 57%; African-Americans, 43%; and Hispanics, 47% (Rainie et al. 2001, p. 2).

Internet surveying has several features that make it commercially attractive and provide strong economic incentives for its development. These features are also likely to be attractive to researchers.

First, Internet surveying has extremely low marginal costs. Telephone surveys have relatively high marginal costs because they involve the time of interviewers and supervisors. Time costs accumulate in proportion to time spent with respondents and the time spent trying to reach respondents. In contrast, server technology makes the marginal costs of distributing surveys and receiving responses by Internet extremely low. Low marginal costs imply a larger sample size for a given budget, which can then facilitate the comparison of multiple designs within the same sample. Although we do not have information on the marginal costs of sampling from the Harris Interactive or Knowledge Networks panels used in this study, we can provide the following comparison of commercial rates for an 18-min survey: Knowledge Networks, \$60,000 for 2000 completions; Harris Interactive, \$35,000 for 2000 completions and \$72,000 for 6000 completions. For comparison, our telephone survey, with about 1700 completions, cost approximately \$50,000. The first Harris Interactive sample actually cost the project \$40,000; the second Harris Interactive and the Knowledge Networks samples were provided gratis, suggesting relatively low marginal costs.

Second, Internet surveys allow for the provision of more, and more varied, information than telephone surveys. The capability to provide respondents with audiovisual information allows for more representative and systematic evaluations of advertisements and new products than can be obtained from the commonly used focus groups. Tracking respondents' use of information can also provide a basis for assessing the degree of respondent effort (Berrens et al. 2002).

Third, Internet surveying permits rapid collection of data. When surveys are components of product design cycles or political campaigns, the capability to collect data rapidly may permit more frequent consideration of alternative strategies. Rapid data collection also has value in political polling and electoral research, especially in elections in which large numbers of voters make candidate decisions close to the election date. Social scientists interested in the effect of events on short-term public opinion are also likely to find the capacity for rapidly drawing large samples attractive.

Fourth, the low marginal cost of Internet surveying facilitates the identification of respondents with relatively rare characteristics. Social scientists studying a wide range of rare populations, including those with specific combinations of demographic and political attributes, typically face a serious “needle-in-the-haystack” sampling problem. For example, if one were interested in identifying a sample of people who have volunteered in political campaigns to learn more about the motivations for this type of political participation, Internet sampling might be feasible where RDD would be prohibitively expensive.

Of course, the major problem with Internet surveying is sampling. No technology comparable to RDD exists for sampling Internet users. Furthermore, if such technology did exist, it would almost certainly violate prohibitions against spamming. Two methods for dealing with the sampling problem have been developed: large-panel and random-panel assembly.

### 2.1 *Large-Panel Assembly*

Large-panel assembly has been pioneered by Harris Interactive (HI, formerly Harris Black International). The approach involves recruiting Internet users into a panel of willing respondents through various means, including advertisements and sweepstakes, the Harris/Excite poll, telephone surveys, and product registrations on Excite and Netscape (Taylor et al. 2001). Currently, the panel includes about 7 million adults who can be randomly sampled for particular surveys.

From the perspective of traditional survey research methodology, the HI approach seems unlikely to provide representative samples of the U.S. population. Coverage error is obviously a major concern, given that only about one-half of U.S. adults currently have Internet access. In addition, the practice of sending out large numbers of invitations with relatively short periods for response leads to low response rates, and hence raises concerns about nonresponse error. As one survey researcher noted, “At best, we end up with a large sample representing nothing but itself” (Mitofsky 1999, p. 24).

Nevertheless, HI recently had an exceptionally strong showing in one of the few survey applications in which there is an objective measure of performance—election forecasting. From October 30 through November 6, it polled 300,000 adults, processing more than 40,000 interviews per hour. Overall, the Internet poll did better in predicting state-level presidential votes than did the final telephone polls of other firms conducted on or after October 27: for the 38 states in which HI polled by Internet, its polls were off an average of 1.8% for Gore and 2.5% for Bush, whereas the telephone polls were off an average of 3.9% for Gore and 4.4% for Bush (RFL Communications 2000; Rademacher and Smith 2001). The Internet polls also correctly called 26 of 27 Senate races with an average error for the two major candidates of 2.2%, and correctly called seven out of seven governor’s races with an average error for the two major candidates of 1.9% (Taylor et al. 2001, p. 38). Although success in election polling depends on more than just the data collected, the exceptionally strong performance of HI in predicting 2000 election races relative to established telephone polling firms suggests that it has found a way to control survey error.<sup>1</sup>

Many surveys estimate weights to make their samples look more like the population being sampled. Telephone surveys often weight to match known population characteristics, such as age, sex, and regional distributions. In-person surveys must usually estimate weights

---

<sup>1</sup>In addition, in June and July 2000, Krosnick and Chang (2001) administered parallel surveys on opinion and voting preferences for the presidential election using the same three modes (telephone RDD, Harris Interactive, and Knowledge Networks). They concluded: “This study suggests that Internet-based data collection represents a viable approach to conducting representative sample surveys” (p. 7).

to compensate for the necessity of sampling in clusters. Weighting is likely to be especially important for Internet samples, given the differences between the sampling frame and the population. In addition to estimating standard weights to bring sample marginals closer to population marginals in terms of demographic characteristics, HI has applied propensity weighting in an effort to take account of behavioral and attitudinal characteristics that might distinguish Internet samples from telephone samples (on propensity weighting, see Rosenbaum and Rubin 1984; Rubin 1997; D'Agostino and Rubin 2000).

The propensity weighting involves adding attitudinal and behavioral questions to RDD telephone and Internet surveys being conducted contemporaneously, although typically for different purposes. The telephone and Internet data are merged and the attitudinal questions and standard demographic variables are used to predict the probability of being in one sample rather than the other. These probabilities, or propensities, then serve as the basis for weighting the Internet sample so that its pattern of covariates, including the attitudinal and behavioral questions, match those in the telephone sample.

For example, consider a study that used an RDD telephone survey with an Internet survey of a panel recruited through one of the methods described in Alvarez et al. (2003). One might expect that people who join panels after visiting a Web page would tend to be more inquisitive about politics than people randomly selected from the population, even after controlling for standard demographic characteristics. Imagine that each survey asked questions about recent activities related to gathering information, such as reading books, consulting friends, or watching television news. Also assume that the two samples do not separate perfectly on these questions. Propensity weighting could be implemented with the following three steps. First, pool the data for the two samples together with a variable indicating from which sample each observation came. Second, estimate a model with the sample indicator as the dependent variable and the information gathering questions and standard demographics as the independent variables, and predict the probability that each observation will be in the telephone sample, which is taken as representative of the population. Third, weight cases in the Internet sample so that probability deciles have the same proportion of cases from the two samples. For instance, if one decile has a smaller proportion of Internet cases than the others, the cases that fall in that decile would be given larger weights.

## 2.2 *Random Panel Assembly*

Formerly known as Intersurvey, Knowledge Networks (KN) has adopted an alternative approach based on random sampling of the general population into a panel of Web TV-enabled respondents. List-assisted RDD is used to identify random samples of households. Efforts are made to recruit the 84% of sampled households located in geographic areas with Web TV ISP Network coverage. Mailing addresses for about 60% of the sampled numbers are identified and advance letters, containing either \$5 or \$10, are sent just prior to telephone contact. Sample numbers are called up to 15 times to reach one adult respondent per household. Recruited households are provided, free of charge, a Web TV unit (an Internet appliance that connects to a telephone and television), Web access, e-mail accounts for all those in the household 13 years and older, and ongoing technical support. The panel members thus take surveys on standardized equipment. Panel members agree to participate in at most one survey of approximately 10–15 min duration per week. Various incentives, including cash and prizes, are intermittently given to households that stay in the panel.

Approximately 56% of contacted households initially agree to join the panel. Of these, 72% allow Web TVs to be installed, and 83% of those complete the core member and core household profiles needed to enter the panel. On average, surveys assigned to panel

members have a response rate of approximately 75%. Taking attrition at each stage into account yields an overall response rate of about 25%. Currently, the panel consists of more than 100,000 members.

Although perhaps not “the most perfect sample of Americans in the history of polling” (Lewis 2000, p. 64), the KN panel has a very strong basis for providing nationally representative samples comparable to those provided by telephone surveys. The coverage and sampling frame are essentially the same as for RDD telephone surveys. Although the overall response rate is probably lower than for the better telephone surveys, this is mitigated to some extent because information known about panel members who do not complete assigned surveys can be used to control statistically for that component of nonresponse error. In terms of measurement error, there is a risk of *panel conditioning*, or *time-in-sample effects*—changes in item responses resulting from the experience of having been previously surveyed. There have been a small number of investigations of economic (Silberstein and Jacobs 1989; Citro and Kalton 1993) and political (Bartels 1999) surveys that have found some evidence of panel conditioning. Although the evidence is limited, it appears that “conditioning effects do sometimes occur, but they are not pervasive” (Kalton and Citro 1993, p. 211). KN currently anticipates keeping participants in the panel for no more than 3 years to reduce the risks of panel conditioning.

### 3 Project Purposes

In addition to assessing public attitudes concerning global climate change, the study design included application of a commonly used method for valuing environmental changes. The contingent valuation (CV) method has become a prominent survey-based approach for valuing nonmarket goods. Statements of willingness-to-pay are elicited from respondents, using various question formats, for proposed changes in public goods or policies.<sup>2</sup> Where there are no observable behavioral traces associated with the public goods, CV may be the only way to value them. CV has been the subject of much methodological debate. In 1993, a blue-ribbon panel of social scientists, convened by the National Oceanic and Atmospheric Administration, further legitimized the use of CV for public policy purposes by concluding that it could be the basis for estimating passive use values in natural resource damage assessment cases (Arrow et al. 1993). Although most applications and methodological research into CV deal with environmental issues, it is seeing increasing use in other areas of public policy in which researchers seek a money metric for public goods.

One of the purposes of this study is to answer several methodological questions through analysis of parallel survey data collected through three modes: an RDD telephone sample, samples from the HI large panel (two waves), and a sample from the KN random panel. First: Could the lower-cost Internet samples produce estimates of willingness-to-pay functions comparable to those from the more expensive telephone surveys? Second: Could splits within the Internet sample be used reasonably to investigate methodological issues? In particular, does the inclusion of questions that encourage respondents to think more carefully about their discretionary income affect willingness-to-pay? Does the provision of extensive information related to the policy affect respondents’ willingness-to-pay? Third: What is the willingness of the U.S. population to pay for ratification of the Kyoto Protocol? These questions are addressed elsewhere; here, we take advantage of a number of questions

---

<sup>2</sup>For overviews of CV and further references, see Mitchell and Carson (1989), Bateman and Willis (2000), and Boardman et al. (2001).

asked of respondents to assess more generally whether Internet surveying has progressed sufficiently to be a viable alternative to telephone surveys in social science research.

The study involves three “treatments” across the survey modes in a  $2 \times 2 \times 2$  design. First, approximately one-half of the respondents in each survey mode were given two “mental accounts” questions that asked them to estimate their disposable income and their contributions to environmental organizations and causes, whereas the others received only the standard CV reminder that payments for the public good would come at the expense of other items in their budgets. Second, about one-half of the Internet respondents were given access to “enhanced information” (27 one-page entries) about the science of global climate change and the Kyoto Protocol, whereas the rest received only descriptive information about the Kyoto Protocol. Third, about one-half of the Internet respondents were given a referendum question on the actual Kyoto Protocol, whereas the others were given a referendum question on a version of the Kyoto Protocol modified to include mandatory emissions limits on greenhouse gases for developing countries.

#### 4 Survey Instrument

The survey instrument had three major sections.<sup>3</sup> The first asked questions on demographic information, environmental attitudes, and knowledge about global climate change and the Kyoto Protocol. The next section implemented the mental accounts and enhanced information treatments. The section then asked questions related to household willingness-to-pay for Senate ratification of the Kyoto Protocol, or the modified Kyoto Protocol, including how respondents would vote in an advisory referendum for their senators if ratification would cost their households a specified annual dollar amount in higher taxes and energy prices. The dollar amount, or “bid” price, was drawn with equal probability from the following list: 6, 12, 25, 75, 150, 225, 300, 500, 700, 900, 1200, 1800, and 2400. The final section asked questions about the fairness of making public policy decisions on the basis of willingness-to-pay, political attitudes and participation, additional demographic data, and, for those given access to the enhanced information on global climate and the Kyoto Protocol, their perceptions of the usefulness and fairness of the information. The survey can be viewed at <http://www.unm.edu/instpp/gcc/>.<sup>4</sup> With few exceptions, the wording and order of questions in the telephone script were replicated exactly in the Internet instrument. Several attitudinal questions were added to the end of the second HI sample to facilitate propensity weighting; the KN sample included some standard proprietary questions at the end.

#### 5 Data from Parallel Surveys

The project as originally funded by the National Science Foundation called for an RDD national telephone sample of approximate 1200 completed surveys to be collected by the

<sup>3</sup>Development of the survey instrument began in 1998 as part of the preparation of a grant application to the National Science Foundation. On short notice, HI generated an Internet sample ( $n = 869$ ) to provide comparisons with questions on global climate change that had appeared in a national telephone survey conducted by the Institute of Public Policy at the University of New Mexico in late 1997. After receipt of the grant, a focus group was held at the Institute for Public Policy. A “beta” version Web survey instrument was constructed to help in designing a survey instrument that could be administered by both telephone and Internet. The beta version included the 27 pages of information on global climate change and the Kyoto Protocol developed collaboratively by the authors. A CATI version of the survey was prepared and provided to HI and KN. HI prepared and pretested its survey instrument in December 1999. Implementation of the telephone survey began prior to administration of the Internet version.

<sup>4</sup>Visitors to the Web site are randomly assigned to treatments. Those wishing to see specific treatments, such as the enhanced information pages, may thus have to visit the site several times.



Institute for Public Policy at the University of New Mexico, and a contemporaneous Internet sample of 6000 to be collected by HI from its panel of willing respondents. Subsequently, HI provided a replication gratis and KN provided a sample from its panel gratis. Consequently, four samples are available for comparison.

The telephone sample, including an initial pretest, was collected between November 23, 1999, and January 27, 2000. It was drawn using an RDD frame, with nonworking numbers stripped, that was purchased from Survey Sampling, Inc., of Fairfield, Connecticut. The surveys were administered by weekday evening and weekend shifts using a 19-station CATI laboratory at the Institute for Public Policy. Sampled numbers were called up to 12 times before being abandoned; hanging appointments were called up to 20 times. Surveys took approximately 15 min to complete on average. The yield was 1699 completed surveys. The response rate was 45.6% based on AAPOR (1998).<sup>5</sup>

Probability weights were constructed for two purposes. First, the weights were inversely proportional to the number of telephone lines of the household to take account of over-sampling as a result of multiple telephone numbers. Second, the weights were proportional to the number of adults in the household to facilitate comparison with samples of individuals.

In January 2000, HI sent invitations to participate in the study to a random sample of its panel of 4.4 million willing U.S. adults. Those invited were given the address of a Web page containing the survey and a password to access it. Those beginning the survey could exit and reenter the Web page until they completed the survey. The survey was closed shortly after quotas for all of the survey splits were obtained. The total yield was 13,034 completed surveys collected between January 11 and 19. The response rate, calculated as the ratio of completed surveys to invitations sent, was 4.0%.

In order to weight the sample to match the demographics of U.S. adults better, HI used a raking procedure (Deming and Stephan 1940; Zieschang 1990; Deville et al. 1993). The weights were selected to match 32 known demographic marginals: four age groups, four regions, and sex. Subsequently, the same procedure was applied to the telephone sample to create a second set of telephone weights for comparison purposes.

In July 2000, HI invited a random sample of its 4.8 million willing U.S. adult respondents to participate in a replication of the survey, yielding a sample size of 11,160 collected between July 10 and 17. The response rate, based on invitations sent and completed surveys, was 5.5%. This sample was propensity-weighted based on attitudinal and behavioral questions concurrently being asked in HI RDD telephone surveys.<sup>6</sup>

From November 25, 2000, to December 11, 2000, KN administered the survey to a random sample of its panel based on previously estimated probability weights to correct for nonresponses in the selection stages in the panel. Only one respondent was selected per household. Of those sampled, 76% completed surveys, yielding a sample size of 2162 and a multistage response rate of 24.1% (20.2% taking account of Web TV noncoverage). For this analysis, raking weights based on the Current Population Survey were estimated for the sample with respect to age, gender, race, ethnicity, region, education, and metropolitan

---

<sup>5</sup>The formula used for the response rate is *completes plus partials* divided by *completes plus partials plus "break offs" plus unfinished appointments plus refusals plus those not interviewed as a result of a language barrier plus those too ill to be surveyed*.

<sup>6</sup>HI uses several different question sets for propensity weighting. In this study, in addition to three attitudinal questions about whether Washington was in touch with the rest of the country, personal efficacy, and information overload, respondents were asked if they owned a retirement account and whether they had read a book, traveled, or participated in a team or individual sport over the past month.

versus nonmetropolitan household location to correct further for nonresponse bias. These weights also convert the data from a household to a population sample.

## 6 Survey Mode Comparisons

In the following sections we present a number of comparisons across the survey modes. In general, the telephone sample is taken as the basis of comparison. Two caveats are worth noting. First, the telephone sample should not be viewed as a perfect sample; it has all the flaws common to RDD telephone samples.<sup>7</sup> The comparisons should be viewed as answering the question: How do the Internet samples compare to a high-quality telephone sample of the sort commonly used in social science research? Second, although all four surveys were collected within a span of 11 months, only the first HI sample is contemporaneous with the telephone sample. Underlying changes in the population cannot be ruled out as explanations for differences between these two samples and the other two.

### 6.1 Socioeconomic

Demographic comparisons across the modes are presented in Table 1. The first two rows show mean age and percentage male. The weighted data (shown in bold) produce fairly close figures for all four samples. The next row, percentage of respondents with at least a college degree, shows considerable difference between the telephone and Internet samples. As is often the case in telephone surveys, the telephone sample overestimates the percentage of adults with a college degree—41.4% as opposed to 23.2% estimated in the March 2000 Current Population Survey. The percentages for the Internet samples are very close to the Census Bureau estimate. Whereas the HI unweighted sample percentages for college degree are gross overestimates, the KN sample percentage is very close, reflecting to some degree the use of probability weights in sampling from its panel.

The three Internet samples slightly underestimate the population percentages of Hispanics (12.5% in the 2000 Census) and African-Americans (12.9% in the 2000 Census), whereas the telephone sample substantially underestimates these percentages. The mean household income in the telephone sample is very close to the population mean of \$57,000 (March 2001 Current Population Survey), but surprisingly, the Internet samples show substantially lower estimates. One striking, but not unexpected, difference is in the percentage of households with a computer—the HI sample percentages are much larger than those in the telephone and KN sample. Looking just at those in the telephone sample who use the Internet at least weekly, the percentage with home computers is much closer to the HI samples.

Given substantial item nonresponse, some caution is needed in interpreting the income figures.<sup>8</sup> The mean household income was largest for the telephone sample, and smallest for the first HI sample. As expected, the HI households, with universal Internet use and high rates of home computer ownership, have substantially larger mean numbers of telephone lines than do either the telephone or KN samples. The telephone sample substantially overestimates the percentage of the population with college degrees and underestimates the African-American and Hispanic percentages.

<sup>7</sup>The Institute for Public Policy has been conducting RDD polls since the early 1990s. Its surveys provide data for studies published in a variety of social science journals.

<sup>8</sup>The item response rates for income were as follows: telephone, 84.9%; first HI, 79.8%; second HI, 82.1%; and KN, 70.9%.

**Table 1** Comparison of respondent socioeconomic characteristics across surveys

Survey mean (standard error) [population]	Public Policy Institute January telephone (n = 1699)				Harris Interactive January Internet (n = 13,034)		Harris Interactive July Internet (n = 11,160)		Knowledge Networks November Internet (n = 2162)	
	Household weighted <sup>a</sup>				Raking weighted <sup>b</sup>		Raking weighted <sup>b</sup>		Raking weighted <sup>d</sup>	
	Full sample	No (N = 726)	Yes (N = 973)	Raking weighted <sup>b</sup>	Raw	Raking weighted <sup>b</sup>	Raw	Propensity weighted <sup>c</sup>	Raw	Raking weighted <sup>d</sup>
Mean age in years [45.7]	42.6 (.42)	46.8 (.68)	39.3 (.49)	44.7 (.48)	41.6 (.10)	44.4 (.71)	42.6 (.13)	44.4 (.50)	45.8 (.36)	44.6 (.42)
Percentage male [49.1%]	47.6 (1.3)	42.5 (2.0)	51.7 (1.8)	47.9 (1.3)	44.3 (.44)	48.0 (1.4)	56.7 (.47)	48.0 (1.3)	49.4 (1.1)	47.8 (1.2)
Percentage college graduate [23.2%]	41.4 (1.3)	26.5 (1.8)	53.4 (1.8)	42.7 (1.3)	43.5 (1.2)	22.0 (.71)	45.9 (.47)	22.9 (.79)	23.9 (.92)	21.2 (.94)
Percentage Hispanic [12.5%]	6.9 (.74)	6.8 (1.1)	6.9 (.99)	10.0 (.97)	3.1 (.15)	9.4 (.96)	2.9 (.16)	9.7 (1.1)	9.8 (.63)	10.4 (.76)
Percentage African-American [12.9%] <sup>e</sup>	7.6 (.71)	8.7 (1.1)	6.7 (.89)	12.9 (1.1)	3.0 (.15)	12.4 (1.3)	2.7 (.15)	11.5 (1.1)	9.3 (.63)	10.8 (.82)
Household mean income (in \$1000s) [57.0%]	56.5 (1.2)	44.8 (1.6)	65.8 (1.6)	57.4 (1.4)	51.3 (.34)	45.1 (1.6)	55.7 (.40)	52.2 (1.2)	49.4 (.84)	46.3 (.85)
Percentage with computers at home	64.5 (1.3)	37.0 (2.0)	86.4 (1.3)	62.7 (1.3)	93.5 (.22)	93.0 (.67)	95.3 (.20)	95.9 (.43)	60.9 (1.1)	58.2 (1.3)
Percentage with computer at work	67.2 (1.2)	51.6 (2.0)	80.3 (1.4)	66.7 (1.2)	66.0 (.41)	54.6 (1.4)	66.1 (.45)	50.3 (1.3)	48.4 (1.1)	47.4 (1.2)
Mean number of telephone lines	1.19 (.016)	1.10 (.011)	1.26 (.027)	1.30 (.021)	1.40 (.0058)	1.40 (.023)	1.41 (.0064)	1.38 (.018)	1.10 (.0074)	1.06 (.0049)

<sup>a</sup>Weights proportional to adults in households divided by number of telephone lines to convert from household level to individual level.

<sup>b</sup>Weights set to match 32 national marginals: regions (four categories), sex (two categories), and age cohorts (four categories).

<sup>c</sup>Weights based on propensity scores estimated by Harris Interactive using data from parallel telephone surveys.

<sup>d</sup>Weights based matches to know demographic marginals and corrections for sample selection bias.

<sup>e</sup>Percentage black or African-American, or most closely identify with black or African-American if of mixed race.

## 6.2 Environmental Knowledge

Socioeconomic differences among the samples do not necessarily impose a fundamental problem in that statistical adjustments can be made in analyses to take account of the *observable* differences. At the same time, even if the samples were identical in terms of socioeconomic characteristics, they could still produce different inferences about relationships among variables in the population because they differ in terms of *unobservable* characteristics. Although it is never possible to know which unobservable characteristics are relevant to any particular analysis, it is interesting to explore differences across the samples, where possible. Similarity of samples in terms of the knowledge and attitudes that we can measure offers some reassurance that inferences based on them have external validity.

Survey questions intended to elicit respondents' knowledge about scientific views on the likely causes and consequences of global climate change provide a basis for comparison. Table 2 compares the percentage of sample respondents with correct answers to 10 environmental knowledge questions, recognition of the Kyoto Protocol, and an overall knowledge score constructed as the sum of correct answers and recognition of the Kyoto Protocol. When "Don't know" responses are treated as incorrect answers (leftmost column under each mode), the KN sample percentages appear substantially and systematically smaller than those for the telephone or HI samples. When "Don't know" is treated as a missing value (the rightmost columns under each mode), the KN sample percentages are no longer systematically smaller than those for the other modes.<sup>9</sup> Figure 1 displays the correspondence between the percentages of the Internet samples correctly answering each knowledge question and the percentage of telephone respondents answering the question correctly. The large number of pluses that lie below the line, representing equality, are the KN percentages when "Don't know" is taken as an incorrect response.

In order to investigate statistical significance, individual-level probit models for each of the 11 knowledge questions in Table 2 were estimated: the dependent variable was whether the respondent correctly answered the question (1 if yes, 0 if no), and the independent variables were indicator variables for the three Internet samples.<sup>10</sup> The 11-point knowledge score, listed in the last row of Table 2, was modeled as an ordered probit. For any given knowledge question, asterisks indicate statistical significance at the 5% level. The large sample sizes for these estimations mean that they have large power for finding statistically significant differences. Inclusion of demographic variables in the estimations generally did not wash out the mode effects.

The *pattern* of correct responses across the 11 knowledge questions can be investigated using Wilcoxon matched-paired signed-rank tests. When "Don't know" is treated as an incorrect answer, the patterns of responses do not statistically differ between the telephone and HI samples at the 5% level. Substantively, they show relatively small average percentage differences (7.0% and 7.4%). The KN sample differs statistically from all three of the others and shows a large average percentage difference with the telephone survey (24.6%). The picture changes substantially when "Don't know" is treated as a missing value. The KN distribution is no longer statistically different from the telephone sample, but is statistically different from the second HI sample. In addition, although the percentage difference remains

<sup>9</sup>Mondak (1999) argues against the common practice of treating "Don't knows" as incorrect answers in the construction of knowledge scales. His analysis suggests that treating "Don't know" as missing provides a more meaningful comparison.

<sup>10</sup>All statistical estimations treat the modes as survey strata, each with their own set of probability weights. The estimations were done using the STATA Version 7 statistical software package.

**Table 2** Comparison of respondent knowledge across surveys

Survey question	Public Policy Institute January Telephone		Harris Interactive January Internet		Harris Interactive July Internet		Knowledge Networks November Internet	
	“Don’t know” as incorrect	“Don’t know” as missing	“Don’t know” as incorrect	“Don’t know” as missing	“Don’t know” as incorrect	“Don’t know” as missing	“Don’t know” as incorrect	“Don’t know” as missing
E <sup>a</sup> : Temperature rises (%Y)	89.3	94.9	87.4	96.0	88.8	96.0	75.9*	95.0
E: Ocean levels fall (%N)	52.4	63.4	45.8*	62.4	40.5*	56.1*	33.1*	54.1*
E: More droughts (%Y)	75.6	85.6	74.3	90.2*	77.6	92.4*	61.7*	90.7*
E: Fewer floods (%N)	68.4	80.8	63.4*	88.3*	63.5*	87.8*	46.5*	83.1
E: More storms (%Y)	85.4	92.9	84.3	95.2*	83.0	93.4	70.2*	94.4
C <sup>b</sup> : Exhaust (%Y)	87.2	92.5	88.4	94.2	89.4	95.9*	78.2*	96.6*
C: Nuclear (%N)	32.2	41.1	28.9	42.4	28.8*	43.6	17.1*	29.1*
C: Toxics (%N)	31.9	39.7	23.8*	32.2*	27.2*	38.3	15.8*	24.7*
C: Coal (%Y)	53.1	70.2	57.0*	81.2*	58.6*	85.8*	50.1	85.3*
C: Forest loss (%Y)	83.8	90.4	86.1	93.6*	86.0	94.4*	75.3*	95.6*
H <sup>c</sup> : Heard of treaty (%Y)	14.4	14.5	15.8	15.8	14.5	14.5	10.5*	10.5*
Knowledge score (0–11)	6.74	7.14	6.55*	7.37*	6.58	7.54*	5.34*	7.14

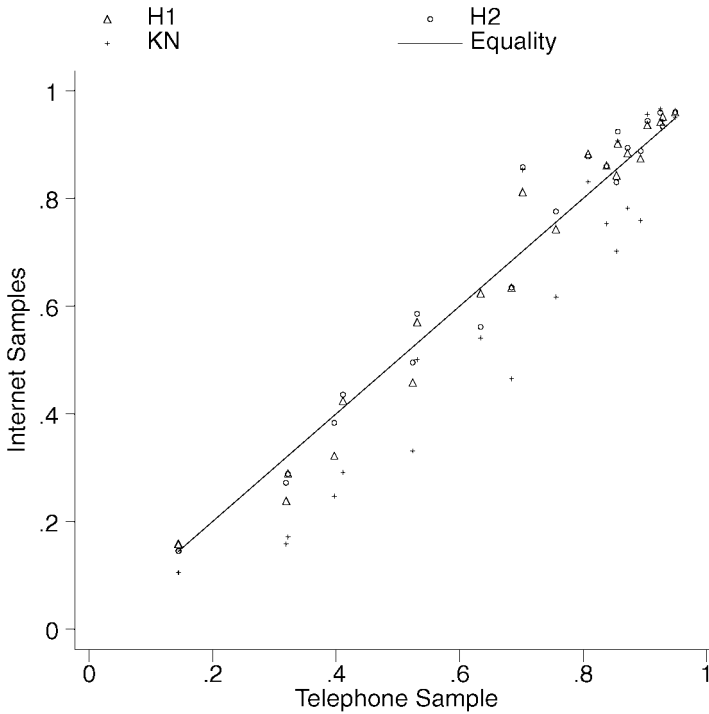
Note. Telephone data weighted to individuals; Internet surveys with proprietary weights.

\*Indicates a statistically significant mode effects (relative to telephone mode) in probit regressions on individual level data. Eleven items based on dichotomous probits; knowledge score based on ordered probit.

<sup>a</sup>Effects (E): Scientists who specialize in the study of the Earth’s climate have debated the possible effects of climate change. Do most scientists expect any of the following changes in global climate to take place? Do most scientists expect...?

<sup>b</sup>Causes (C): Many scientists have argued that global average temperatures have risen slightly and will continue to increase for many years as a result of human activities. To the best of your knowledge: Do scientists believe...?

<sup>c</sup>Treaty (K): Have you heard about the proposed international treaty called the Kyoto Protocol?



**Fig. 1** Scatter plot of percentage correct.

small (7.2%), the distribution of the telephone sample is statistically different from the distribution of the first HI sample.

Overall, there appear to be statistically significant differences in environmental knowledge among the survey modes, but these differences generally appear to be substantively small. The higher rates of “Don’t know” in the KN sample could possibly be an indication of panel conditioning—either fatigue or changing norms of response.<sup>11</sup>

### 6.3 Information Use in Internet Modes

The use of access to enhanced information offered to subsamples of Internet respondents provides an opportunity for comparing survey motivation among the HI and KN samples. Use rates were very similar across the three samples: 72.7%, 68.8%, and 66.2%, for the January HI, July HI, and KN samples, respectively. On average HI respondents reported viewing more pages (7.1 and 5.5) than KN respondents (3.8). The average reported time spent viewing pages was very similar across the samples: 9.4 min in each of the HI samples and 9.0 min in the KN sample. Perceptions of usefulness of the information (0, not at all useful; 10, extremely useful) and its perceived bias

<sup>11</sup>Only the number of previous surveys completed are available in the KN data set. The number of previous completions does not appear to have any significant effect on the total number of “Don’t knows” in the eleven-question set for males. There appears to be a weak quadratic relationship between “Don’t knows” and previous completions for females—suggesting that “Don’t knows” fall during the first 14 completions and rise thereafter. In the sample, the mean number of previous completions was 18.

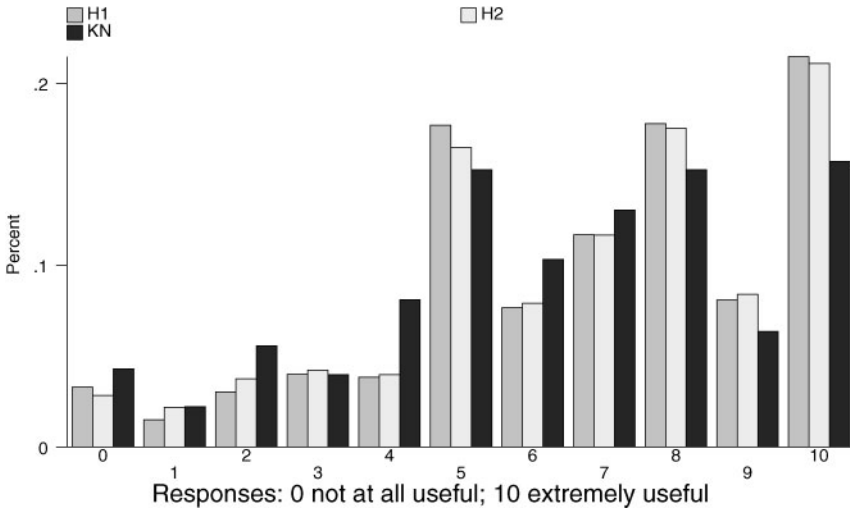


Fig. 2 Usefulness of information.

(0, strongly against GCC; 10, strongly in favor of GCC) were similar across the samples. The means for usefulness were 7.0 for each of the HI samples and 6.5 for the KN sample; the means for bias were 5.9 for each of the HI samples and 5.6 for the KN sample.

Do the distributions of responses to the usefulness and bias questions show similar patterns across the Internet samples? Figures 2 and 3 (based on unweighted data) display response frequencies for these two evaluative questions. The three samples show roughly similar patterns. Overall, information users in the Internet samples appear to have perceived the information they accessed in roughly the same way.

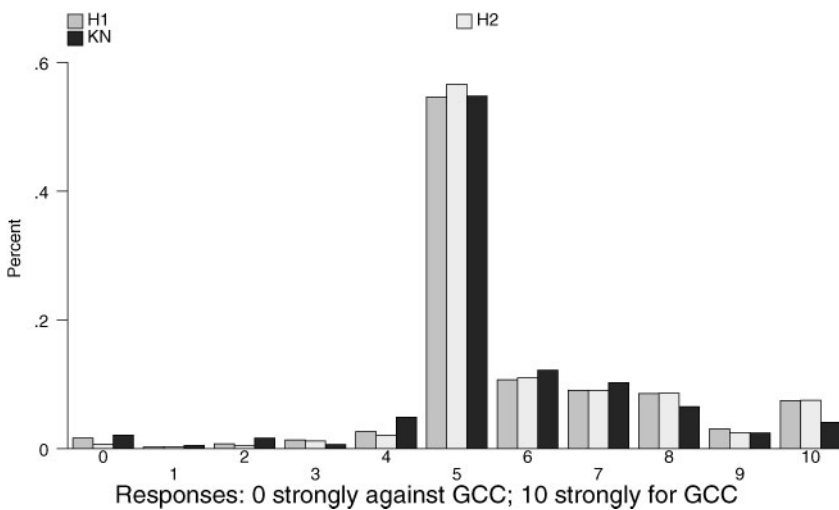


Fig. 3 Bias in information.

**Table 3** Comparison of political variables and environmental attitudes across surveys

<i>Survey question</i>	<i>January Public Policy Institute Telephone Household weighted</i>	<i>January Harris Internet</i>		<i>July Harris Internet</i>		<i>November Knowledge Networks</i>	
		<i>Raw</i>	<i>Raking weighted</i>	<i>Raw</i>	<i>Propensity weighted</i>	<i>Raw</i>	<i>Raking weighted</i>
Registered to vote (NES 2000: 85%) <sup>a</sup>	<b>86.7</b>	89.5	<b>84.5</b>	91.2	<b>87.4</b>	76.6	<b>72.9</b>
Democrat (NES 2000: 50%)	<b>34.4</b>	31.6	<b>36.8</b>	28.5	<b>37.5</b>	40.6	<b>41.5</b>
Republican (NES 2000: 37%)	<b>33.9</b>	28.4	<b>24.1</b>	33.1	<b>32.3</b>	29.6	<b>27.7</b>
Third party (%)	<b>2.8</b>	5.1	<b>4.1</b>	5.5	<b>2.8</b>	3.9	<b>3.6</b>
Members of environmental groups (%)	<b>10.9</b>	16.3	<b>11.8</b>	15.1	<b>9.5</b>	6.5	<b>6.4</b>
Ideology (7-Point scale; 1, strongly liberal) (NES 2000: 4.2)	<b>4.29</b>	4.06	<b>4.03</b>	4.21	<b>4.11</b>	4.09	<b>4.04</b>
Environmental threat (11-point scale; 0, no real threat; 10, brink of collapse)	<b>5.71</b>	5.86	<b>5.83</b>	5.72	<b>5.74</b>	5.42	<b>5.48</b>
International environmental treaties (11-point scale; 0, very bad idea; 10, very good idea)	<b>7.20</b>	6.9	<b>6.93</b>	6.69	<b>6.78</b>	6.87	<b>6.82</b>
Property rights over environmental protection (4-point scale; 1, strongly disagree)	<b>2.66</b>	2.44	<b>2.53</b>	2.52	<b>2.59</b>	2.53	<b>2.53</b>
Ideology: percentage 1 (percentage 7)	<b>4.4</b> <b>(6.7)</b>	4.9 (6.5)	<b>5.1</b> <b>(6.7)</b>	4.8 (7.5)	<b>5.0</b> <b>(5.7)</b>	4.9 (6.0)	<b>4.7</b> <b>(6.3)</b>
Environmental threat: percentage 0 (percentage 10)	<b>1.8</b> <b>(2.9)</b>	4.9 (6.5)	<b>4.3</b> <b>(4.9)</b>	4.8 (7.5)	<b>3.7</b> <b>(5.1)</b>	4.9 (6.0)	<b>3.4</b> <b>(4.0)</b>
International treaties: percentage 0 (percentage 10)	<b>4.4</b> <b>(29.6)</b>	6.3 (27.9)	<b>7.0</b> <b>(29.8)</b>	7.2 (25.6)	<b>7.0</b> <b>(27.7)</b>	3.8 (26.1)	<b>4.0</b> <b>(26.3)</b>
Property rights: percentage 1 (percentage 4)	<b>6.8</b> <b>(12.4)</b>	10.2 (11.6)	<b>8.3</b> <b>(14.1)</b>	8.8 (13.2)	<b>7.0</b> <b>(13.3)</b>	7.5 (10.4)	<b>8.0</b> <b>(10.8)</b>

<sup>a</sup>The NES Guide to Public Opinion ([www.umich.edu/~nes/nesguide/gd.index.htm](http://www.umich.edu/~nes/nesguide/gd.index.htm)).



#### 6.4 *Political Variables and Environmental Attitudes*

Of particular interest to political scientists is the comparability of the samples with respect to political attitudes and behavior. Table 3 compares the samples in terms of a number of politically relevant variables. A number of differences appear. The KN sample has a lower rate of voter registration than the others, or the 85% reported by the NES for 2000. It also seems to have a substantially lower rate of membership in environmental groups. All three of the Internet samples seem to be more liberal and have higher fractions of identification with the Democratic party than the telephone sample. (All four samples report lower identification with the two major parties than the NES, although the latter includes leaners; the NES estimate of Ideology is very close to the survey estimates.) The first HI sample has a lower percentage of Republican party identifiers and a higher percentage of third-party identifiers.

The last four rows display the percentages of extreme values for each of the ideology and the three environmental attitudes. Note that the three Internet samples appear to have substantively similar values. The extreme category rates for the telephone sample appear substantively smaller than those of the Internet samples for both extremes of Environmental Threat, and the low extremes for International Treaties and Property Rights. With the large sample size, however, a chi-square test would reject the hypothesis of equal probabilities across the three categories (low extreme, middle, and high extreme) even for Ideology. Thus, it would appear that Internet respondents are more willing to report extreme positions, although the effect appears substantively small.

#### 6.5 *Relationships Between Environmental Views and Ideology*

Although making estimates of population parameters is often important in social science research, much empirical work is directed at testing hypotheses about the relationships among variables. Only when analyses are based on probability samples can we be confident about their generalization to the larger population. As the representativeness of at least the large panel Internet samples is questionable, it is interesting to ask how inferences might differ across modes. We investigate the following general hypothesis: *Political ideology affects environmental attitudes*. Specifically, we investigate the relationship between ideology and: (1) perceptions of environmental threat, (2) tradeoffs between property rights and the environment, and (3) reliance on international environmental treaties.

Table 4 shows the effect of ideology on perception of environmental threat (11-point scale) as estimated in three ordered probit specifications.<sup>12</sup> In the first specification, ideology and its interaction with each of the three Internet modes are the explanatory variables (telephone survey mode as the base category). There are large negative and significant coefficients for ideology under all four of the survey modes. The small and insignificant coefficient for the ideology–KN interaction indicates that we would reach the same conclusion using either sample. The interaction terms for the HI samples show significant impacts of ideology that are about 50% larger than in the other two samples.

As shown in the second column, however, the introduction of a set of standard covariates reduces the size of the coefficients on the interaction terms for the HI samples and washes out their statistical significance. As there was substantial item nonresponse for

---

<sup>12</sup>Results would be qualitatively the same if linear regression rather than ordered probit models were estimated. The results would not hold if the analyses were done using unweighted data—demographic controls generally do not wash out the mode interactions when the data are not weighted.

**Table 4** Effects of ideology on environmental attitudes: crisis?

	<i>No controls</i> ( <i>n</i> = 25,393)	<i>Demographic</i> <i>controls</i> ( <i>n</i> = 20,932)	<i>Demographic controls</i> <i>w/o income</i> ( <i>n</i> = 25,377)
Ideology	-.15* (.018)	-.15* (.020)	-.15* (.018)
Ideology-Internet (H1) Interaction	-.07* (.027)	-.04 (.030)	-.06* (.026)
Ideology-Internet (H2) Interaction	-.07* (.24)	-.05 (.027)	-.06* (.024)
Ideology-Internet (KN) Interaction	-.02 (.025)	-.02 (.030)	-.01 (.11)
Internet (H1)	.33* (.12)	.20 (.11)	.27* (.11)
Internet (H2)	.25* (.11)	.13 (.12)	.20 (.11)
Internet (KN)	-.08 (.11)	-.08 (.13)	-.09 (.11)
F	62.79	31.89	34.31
N	25,393	20,932	25,377
Wald test: equality of H1 and KN interactions with ideology	Not reject	Not reject	Not reject
Wald test: equality of H2 and KN interactions with ideology	Reject	Not reject	Not reject

\*Significant at 5% level.

Dependent variable: Environmental threat (11-point scale; 0, no real threat; 10, brink of collapse).

Ordered probit (cut-points not shown).

Controls: income, age, sex, African-American, Hispanic, college, student, retired, full-time employee, part-time employee, self-employed, homemaker.

Question: Some people believe that pollution, population growth, resource depletion, and other human-made problems have put us on the brink of environmental crisis that will make it impossible for humans to continue to survive as we have in the past. Others believe that these fears are overstated and that we are not in a serious environmental crisis. On a scale from 0 to 10 where 0 means that there is no real environmental crisis and 10 means that human civilization is on the brink of collapse as a result of environmental threats, what do you think about the current environmental situation?

Responses: no real threat (2.7%), 1 (2.3%), 2 (4.1%), 3 (6.7%), 4 (7.8%), 5 (19.4%), 6 (15.6%), 7 (18.8%), 8 (14.9%), 9 (4.31%), brink of collapse (3.5%).

income, the third column shows the model estimated with all demographic covarites except income. The ideology interactions for HI do not lose significance, but they are statistically indistinguishable from the ideology interaction for the KN sample. Nevertheless, across all modes we find a large negative significant relationship between ideology and perception of environment threat.

The same basic patterns result when ideology is used to explain attitudes toward international environmental treaties and tradeoffs between property rights and the environment. In these models, even without controls, both the January HI and the KN mode interactions with ideology have insignificant coefficients, whereas the July HI mode interaction with ideology has a statistically significant coefficient. Adding demographic controls including income makes the July HI mode interaction coefficient insignificant, but the demographic controls without it do not.

It is worth noting that although substantive differences remain relatively small, the demographic controls do not wash out statistically significant differences between the

unweighted HI samples and the telephone and KN samples. Consequently, weighting appears to be important in terms of inference.

To summarize, in these applications, researchers would not make different statistical inferences using either the telephone or the KN samples. Furthermore, if one included income and other demographic controls in the estimation models, one would not make different statistical inferences using the telephone, the KN, or either of the HI samples.

### 6.6 *Referendum Voting Models*

As a final comparison, we investigate mode effects in the basic referendum voting model that underlies CV analysis. We exclude respondents in the Internet studies who were either given access to enhanced information or asked to value the modified Kyoto Protocol, because these treatments did not occur in the telephone sample. The mental accounts treatment, which asked respondents to estimate the percentage of monthly income that was available for discretionary spending and how much of that discretionary income goes toward environmental causes and organizations, was included in all samples.

The “elicitation method” for obtaining information about valuation from respondents used in this study was the advisory referendum format (Carson et al. 1999). After going through a series of questions that were used as vehicles to explain the provisions and likely consequences of ratification of the Kyoto Protocol, respondents were asked the following question:

The U.S. Senate has not yet voted on whether to ratify the Kyoto Protocol. If the U.S. does not ratify the treaty, it is very unlikely that the Protocol can be successfully implemented. Suppose that a national vote or referendum were held today in which U.S. residents could vote to advise their senators whether to support or oppose ratifying the Kyoto Protocol. If U.S. compliance with the treaty would cost your household  $X$  dollars per year in increased energy and gasoline prices, would you vote for or against having your senators support ratification of the Kyoto Protocol? Keep in mind that the  $X$  dollars spent on increased energy and gasoline prices could not be spent on other things, such as other household expenses, charities, groceries, or car payments.  
( $X$  is randomly chosen from: 6, 12, 25, 75, 150, 225, 300, 500, 700, 900, 1200, 1800, 2400.)

In this case, we consider the simplest possible model: a logistic regression with the response to the vote question as the dependent variable (yes = 1, no = 0) and the bid price, income, an indicator for the mental accounts treatment, an interaction between the mental accounts indicator and bid price ( $X$ ), and, in some models, basic demographic controls, as the explanatory variables. If the focus of the analysis were on the estimation of willingness-to-pay (e.g., Cameron and James 1987), then many additional variables would be included and estimation would involve more complicated models. Nevertheless, this simple model, which is representative of the type typically estimated in CV studies as an initial check to see if the data meet minimal validity requirements, allows us to focus on mode effects.

The first column of Table 5 shows the basic model without demographic controls. It shows similar patterns of coefficients as the models for the individual modes shown in the last four columns. Bid price and income have the expected signs and statistical significance; the significant coefficients for the mental accounts indicator and its interaction with bid price shows that the mental accounts treatment reduces the probability of voting yes for bid amounts up to about \$1375, which is near the upper extreme of the bid range. Thus, asking respondents to answer prior questions about their discretionary income generally lowers their probability of voting yes for the referendum. The negative and significant coefficients for the Internet samples indicate that Internet respondents are less likely to vote yes. The first HI sample shows a relatively small effect for which the statistical significance washes out with the addition of demographic controls (Column 2). The second HI and the KN samples

**Table 5** Logistic models of advisory vote for ratification

	<i>All Modes</i> <sup>a</sup>		<i>Telephone</i> ( <i>n</i> = 1358)	<i>Harris</i> <i>Interactive 1</i> ( <i>n</i> = 3186)	<i>Harris</i> <i>Interactive 2</i> ( <i>n</i> = 2451)	<i>Knowledge</i> <i>Networks</i> ( <i>n</i> = 759)
	( <i>n</i> = 7754)	( <i>n</i> = 7748)				
Bid price (in \$1000s)	-.82* (.092)	-.83* (.092)	-.88* (.13)	-.80* (.14)	-.83* (.20)	-.85* (.19)
Income (in \$1000s)	.0027* (.0012)	.0026* (.0012)	.0031* (.0017)	.0018 (.0020)	.0029 (.0025)	.0070* (.0026)
Mental accounts (1, yes; 0, no)	-.42* (.12)	-.43* (.12)	-.29 (.17)	-.20 (.21)	-.72* (.25)	-.44* (.22)
Mental accounts— bid price interaction	.32* (.14)	-.30* (.14)	.079 (.19)	.084 (.21)	.65* (.27)	.36 (.27)
Harris Interactive 1	-.20* (.099)	-.15 (.096)	—	—	—	—
Harris Interactive 2	-.43* (.12)	-.36* (.11)	—	—	—	—
Knowledge Networks	-.40* (.11)	-.35* (.11)	—	—	—	—
Constant	.93* (.12)	1.7* (.27)	.96* (.15)	.73* (.17)	.54* (.21)	.37* (.20)
Demographic controls <sup>b</sup>	No	Yes	No	No	No	No
Adjusted Wald Test	30.6	10.3	21.7	14.3	5.9	8.6
Percentage Yes votes in sample	56.6	56.6	61.1	57.3	55.0	51.0
Percentage correct predictions	61.4	64.6	64.5	61.4	56.4	60.9

\*Significant at the 5% level (income and bid coefficients based on one-sided tests).

<sup>a</sup>The addition of full sets of mode interactions with bid price, income, mental accounts, and the bid price-mental accounts interaction results in no statistically significant mode interaction terms in either model. There were no significant Adjusted Wald tests for particular sets of interactions (i.e. bid price interacted with Harris Interactive 1, Harris Interactive 2, and Knowledge Networks).

<sup>b</sup>Demographic controls: age, sex, African-American, Hispanic, college, student, retired, full-time employee, part-time self-employed, homemaker.

have roughly the same size and remain significant with the addition of the demographic controls.

When the model in the first column of Table 5 is saturated with mode interactions for bid price, income, mental accounts indicator, and the mental accounts–bid price interaction (not shown), the only significant mode effect is the constant shift for the KN sample, which cannot be distinguished statistically from the shift effect for the second HI sample. None of the adjusted Wald tests for the interaction triplets being simultaneously zero can be rejected. With the exception of a lower acceptance rate for the KN sample, there are no consistent mode effects in the referendum model. Furthermore, across all four samples, the analyst would make the same policy inference for the validity test—the probability of voting yes on the referendum is significantly and inversely related to the bid price (or cost) of the policy.

## 7 Conclusion

All survey methods involve errors. The appropriate question is not: Can the Internet replace the telephone as the primary mode of administration in social science survey research? Rather, it is: Under what circumstances is use of Internet surveys appropriate? We have explored this question by making a variety of inferential comparisons of a standard RDD telephone sample with samples from the leading firm in the development of a large panel of willing Internet users (Harris Interactive) and the leading firm in the development of a random panel of Web TV-enabled respondents (Knowledge Networks). Although many differences arose, across a variety of tests on attitudes and voting intentions, the Internet samples produced relational inferences quite similar to the telephone sample. Readers must judge if the similarity we found gives them sufficient confidence to use Internet samples for their particular research questions.

At the same time, Internet surveys based on either large panels or random panels offer possibilities for research that were previously prohibitively expensive. One possibility is the generation of large sample sizes to permit the investigation of methodological questions within the context of the same survey—the large HI sample sizes that allowed us to use a three-treatment design make this point. A second possibility is the opportunity to provide much more information to respondents than is feasible in any other survey mode. Both Internet firms were able to support our enhanced information treatment, and KN was also able to track actual visits to and time spent on particular information pages. A third possibility, not explored in this study, is the capability to generate samples of the population with rare characteristics. Finally, the extension of the HI panel to include willing respondents from other countries opens up intriguing possibilities for comparative analysis of elite attitudes.

## References

- AAPOR 1998. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys*. American Association for Public Opinion Research.
- Alvarez, R. Michael, Robert P. Sherman, and Carla VanBeselaere. 2003. "Subject Acquisition for Web-Based Surveys." *Political Analysis* 11:23–43.
- Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Schulman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(10):4601–4614.
- Arostic, B. K., Nancy Bates, Geraldine Burt, Adriana Silberstein, and Franklin Winters. 1999. "Nonresponse in U.S. Government Household Surveys: Consistent Measures and New Insights." Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, October 28–31.
- Bartels, Larry M. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8:1–20.
- Bateman, Ian J., and Ken G. Willis (eds.). 2000. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the U.S., EC, and Developing Countries*. Oxford: Oxford University Press.
- Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. 2002. "Information and Effort in Contingent Valuation Surveys: Application to Global Climate Change Using National Internet Samples." Manuscript.
- Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. 2001. *Cost–Benefit Analysis: Concepts and Practice*. Upper Saddle River, NJ: Prentice Hall.
- Cameron, Trudy Ann, and Michelle D. James. 1987. "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys." *Review of Economics and Statistics* 69(2):269–276.
- Carson, Richard T., Theodore Groves, and Mark J. Machina. 1999. "Incentives and Informational Properties of Preference Questions." Plenary Address, European Association of Resource and Environmental Economists, Oslo, Norway, June.
- Citro, Constance F., and Graham Kalton (eds.). 1993. *The Future of the Survey of Income and Program Participation*. Washington, DC: National Academy Press.
- Couper, Nick P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464–494.

- CTIA. 2000. *Wireless Industry Indices: 1985–1999*. Cellular Telecommunications Industry Association.
- D'Agostino, Ralph B., Jr., and Donald B. Rubin. 2000. "Estimating and Using Propensity Scores with Partially Missing Data." *Journal of the American Statistical Association* 95:749–759.
- de Leeuw, Edith D. 1999. "Preface." *Journal of Official Statistics* 15(2):127–128.
- Deming, W. Edwards, and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *Annals of Mathematical Statistics* 11(4):427–444.
- Deville, Jean-Claude, Carl-Erik Sarndal, and Oliver Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88:1013–1020.
- eMarketer. 2000. *The eDemographics and Usage Patterns Report*. eMarketer, Inc., New York, September.
- Kalton, Graham, and Constance F. Citro. 1993. "Panel Surveys: Adding the Fourth Dimension." *Survey Methodology* 19(2):205–215.
- Krosnick, Jon A., and Lin Chiat Chang. 2001. "A Comparison of the Random Digit Dialing Telephone Survey Methodology with Internet Survey Methodology as Implemented by Knowledge Networks and Harris Interactive." Ohio State University, April.
- Lewis, Michael. 2000. "The Two-Bucks-a-Minute Democracy." *New York Times Magazine* November 5:64–67.
- Mitchell, Robert C., and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.
- Mitofsky, Warren J. 1999. "Pollsters.com." *Public Perspective* June/July:24–26.
- Mondak, Jeffery J. 1999. "Reconsidering the Measurement of Political Knowledge." *Political Analysis* 8:57–82.
- Piekarski, Linda. 1999. "Telephony and Telephone Sampling: The Dynamics of Change." Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, October 28–31.
- Rademacher, Eric W., and Andrew E. Smith. 2001. "Poll Call." *Public Perspective*. March/April:36–37.
- Rainie, Lee, Dan Packel, Susannah Fox, John Horrigan, Amanda Lenhart, Tom Spooner, Oliver Lewis, and Cornelia Carter. 2001. "More On Line, Doing More." The Pew Internet & American Life Project, Washington, DC, February 18.
- RFL Communications. 2000. "Harris Interactive Uses Election 2000 to Prove Its Online MR Efficacy and Accuracy." *Research Business Report* November 1–2.
- Rosenbaum, Paul R., and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:517–524.
- Rubin, Donald B. 1997. "Estimating Causal Effects From Large Data Sets Using Propensity Scores." *Annals of Internal Medicine* 127:757–763.
- Silberstein, Adriana R., and Curtis A. Jacobs. 1989. "Symptoms of Repeated Interview Effects in the Consumer Expenditure Interview Survey." In *Panel Surveys*, eds. Daniel Kasprzyk, Greg Duncan, Graham Kalton, and M. P. Singh. New York: John Wiley & Sons, pp. 289–303.
- Steeh, Charlotte. 1981. "Trends in Nonresponse Rates, 1952–1979." *Public Opinion Quarterly* 45:40–57.
- Steeh, Charlotte, Nicole Kirgis, Brian Cannon, and Jeff DeWitt. 2000. "Are They Really As Bad As They Seem? Nonresponse Rates at the End of the Twentieth Century." Revision of paper presented to the International Conference on Survey Nonresponse, Portland Oregon, October 28–31.
- Taylor, Humphrey, John Brenner, Gary Overmeyer, Jonathan W. Siegel, and George Terhanian. 2001. "Touchdown! Online Polling Scores Big in November 2000." *Public Perspective* March/April:38–39.
- U.S. Bureau of the Census. 1999. *Statistical Abstract of the United States, 119th Ed.* Washington, DC: U.S. Department of Commerce.
- Walsh, Ekaterina, Michael E. Gazala, and Christine Ham. 2000. "The Truth About the Digital Divide." The Forrester Brief, April 11. (Available from [www.forrester.com/ER/Research/Brief/0,1317,9208.FF.htm](http://www.forrester.com/ER/Research/Brief/0,1317,9208.FF.htm).)
- Zieschang, Kimberly D. 1990. "Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey." *Journal of the American Statistical Association* 85:986–1001.