The Opt-in Internet Panel: Survey Mode, Sampling Methodology and the Implications for Political Research

Seth J. Hill
Corresponding Author: sjhill@ucla.edu
James Lo
Lynn Vavreck
John Zaller

University of California, Los Angeles

All survey methodologies have weaknesses, some known and others unknown. Given that there are no "true" targets against which to assess the political marginal distributions about which we care, the decision to use any survey methodology must be met with caution and full disclosure about the strengths and weaknesses of the method – and how it might affect the results. In this paper, we assess the differences among several national surveys in terms of demographics, marginals on political variables, and ideological constraint among respondents at varying levels of political awareness. Our results suggest that Polimetrix's sample matching technique, used in the 2006 Cooperative Congressional Election Study, seems to produce Internet samples that look more similar to existing RDD phone surveys than to multi-stage probability face-to-face surveys, but surprisingly, many of the discernible differences are not large in size. We conclude that a mildly biased but large Internet survey can produce more reliable estimates than an unbiased but small survey (because of the random error due to the small samples). When matters of cost are factored in, the large, biased sample becomes even more appealing to researchers with limited budgets. The question about Internet samples (even matched samples such as those generated by Polimetrix) remains, however, whether the ignorability assumption holds such that the people who take surveys online behave the same way as those who do not – or more precisely– as those who take phone or in-person surveys.

Through most of the survey era, scholars of political persuasion have rarely had surveys large enough to reliably detect many effects of the magnitude likely to occur. For example, a surge in advertising in presidential battle ground states in the last week of the campaign would probably be invisible in the American National Election Studies, which normally interviews only about 150 respondents in these states in the last week of the campaign. In recent years, however, many scholars have begun to use telephone surveys and most recently "opt-in surveys" on the Internet, which can produce much larger samples at affordable rates. For example, we purchased, for a cost of only about $8 a complete, 3,000 Internet interviews in nine media markets having over 60 congressional races in 2006. The ability to conduct such large, targeted studies creates the potential for major advances in the study of political communication.

Yet, before bounding into the brave new world of Internet surveys, scholars are well advised to pause and evaluate the quality of their data. Can an Internet survey of volunteer respondents generate data that is both valid and representative of the general population? That is the question we investigate in this paper. In particular, we compare data in several reputable national surveys with data from a recent Internet survey, the Cooperative Congressional Election Study (CCES) fielded in 2006 by the research firm Polimetrix, Inc. for a group of over 100 political scientists from nearly 40 Universities. The CCES sampling methodology is designed to generate close correspondence between demographic characteristics of its sample and a target population. But the methodology does not attempt to match respondents to target populations on political engagement or partisan commitment.[1] Thus, while the sample may have the correct proportion of, say, low education people, its low education respondents may be more partisan or more informed than low education persons in the population at large.

Information and engagement are especially important in studies of campaign persuasion because they may affect the degree to which respondents are susceptible to political influence.

[1] We are told by Polimetrix that these variables may be used as targets in the future.

Hence, if the CCES sample does contain too few low-information, moderate, non-partisans, scholars searching for campaign effects may falsely accept the null hypotheses of no persuasion effects when the truth is that the sample is missing the subgroups most likely to experience these effects.

Our analysis of the quality of the CCES sample in this paper produces mixed results. Some tests show statistically significant differences between the CCES sample and other national samples, but other tests are unable to reject the null hypothesis of independence between our sample and a sample based on in-person surveys.

Our overall conclusion is that, although the opt-in Internet sample we analyze does have some bias, the bias is probably not so great as to vitiate the gains of inexpensive, large, and targeted samples that Internet technology make possible, especially if the bias can be minimized through future improvements in the sampling methodology. We argue, therefore, for the use of opt-in Internet samples that are attentive to sample quality, and further suggest that all researchers engaged in survey work, regardless of method, evaluate the biases inherent in their samples before wholeheartedly embracing results.

Background

The growth over the last decade in Internet penetration among Americans has narrowed the digital divide such that before long nearly every household in the country will have access to the Internet. The ability to communicate with people at no cost and on their own schedules has prompted the rise of many survey research companies to conduct research online. The savings to researchers are undeniable. In person interviews such as those conducted by the American National Election Study (ANES) cost upwards of $1000 per complete. Telephone surveys of similar length can be conducted for much less than this, but may cost as much as $200 per complete. In contrast, Internet surveys can be conducted for as little as $15 per complete. As the kinds of questions we ask require more power and precision to answer, the tradeoffs between cost and research method become crucial and the Internet is an affordable, flexible, and time-sensitive way to conduct research. The question, however, is whether the difference in mode results in serious consequences for researchers. To complicate matters, the use of Internet surveys requires working with a panel of respondents who have given the research firm permission to email them. This means that the sampling method for online research is necessarily different than the typical probability samples with which researchers commonly work.

The differences in mode of interview and sampling methodology may produce data that cannot be used to answer certain questions in which social scientists are interested. For example, if the kinds of people who opt-in to these panels are different from the kinds who do not, findings that depend on characteristics determinative of these differences cannot be found. This issue is of particular importance for scholars interested in campaign effects. Decades of work showing minimal effects are being reexamined with more and better data, but results indicate that campaign effects may be large within certain subgroups of the population and perhaps absent in others (Zaller 1992; 1996; Johnston et al 2003; Hillygus and Jackman 2003; Hillygus 2005; Freedman, Franz, and Goldstein 2004). To find campaign effects, survey design must account for the need to represent these population subgroups.

The problem of generating a representative sample is not limited to Internet samples. Election Day exit polls over-represent educated voters. Telephone polls suffer from non-uniform and declining response rates among people with varying levels of education and income (Curtin, Presser, & Singer 2005), and evidence suggests that the date of interview can interact with political or national events to increase the probabilities that certain types of people will complete phone surveys on given days (Hill 2006). Even the National Science Foundation funded political surveys like the expensive American National Election Studies have non-response problems of such significance as to require post-stratification weighting.[2] Every survey methodology has biases, and our goal here is to discover whether an Internet sampling methodology like the one Polimetrix uses can be *made* to produce representative samples, even if the current CCES sample misses the mark. In this paper, we compare Pew Research Center telephone surveys, exit polls, the National Annenberg Election Study telephone survey, and the ANES with the CCES. In all cases we use the post-stratification sampling weights provided with the surveys.

The data for CCES are the product of a survey of 38,443 Americans conducted during October and November of 2006. The data we use here are from the Stanford team module (Jackman 2006) and the Polimetrix team module (Vavreck 2006a). Our own UCLA team module (Vavreck 2006b) was designed in conjunction with the Wisconsin team module (Goldstein 2006) to be representative of 9 media markets in the Midwest where the Wisconsin Advertising Project was tracking advertisements aired by candidates. We will not use these Midwest data here. Respondents were selected from the Polimetrix PollingPoint Panel using a proprietary method called sample matching.

We describe here the procedure for creating the CCES samples. The client/researcher provides Polimetrix with a target population; for example, one might want a nationally-representative sample of size N. Polimetrix draws a random sample of size N from the 2004 American Community Study (ACS), conducted by the U.S. Bureau of the Census, which is a probability sample of size 1,194,354 with a response rate of 93.1 percent. For each respondent in the Polimetrix-drawn ACS sample, the closest matching active Polimetrix panelist is selected using a measure of statistical distance on four Census variables -- age, race, gender, and education, plus on imputed values of partisanship and ideology. The matching is based on joint distributions of the six variables.

[2] The pre-election response rate for the 2004 ANES was 66.1%.

Analysis of demographic variables

A natural point to begin analysis is to examine distributions on the demographic variables to which the CCES sample has been matched in the American Community Survey sample. Table 1 provides this comparison, along with equivalent data from the American National Election Studies Survey (ANES) of 2004 and to the National Annenberg Election Study (NAES), a random digit dial telephone survey of 2004. As can be seen, some of the deviations between the three political surveys and the ACS sample are notable. One substantively large deviation is that 3.9 percent of CCES respondents report less than a high school education compared to 16.9 percent in the ACS sample (the ANES comes closest at14.4). In terms of people with advanced degrees, the NAES contains 14.2 percent, while census figures suggest slightly less than 10 percent of the population are in this category (CCES and ANES do better at 8.7 and 9.9 percent respectively).

Table 2 presents summary measures of deviation from Census demographics. For example, the cell in the upper left of the table states that the mean absolute deviation of each of the four education categories between the aggregated Annenberg totals and the Census totals is 6.1 percentage points. Within each row, the cells are colored with red indicating the survey farthest from the Census, orange indicating the survey second farthest from the Census, and white indicating the survey closest to the Census. The first row indicates that the CCES sample had the largest mean absolute deviation from the Census on education categories while the weighted ANES had the lowest absolute deviation. This is also true for a more standardized deviation measure, the ratio of mean absolute deviation to the Census value, and for a χ^2 test for independence of each column from the Census value.

Subsequent sets of rows present deviation statistics for the other deviation categories. Annenberg does especially poorly on race and gender, while the CCES is less representative in terms of age. The final rows present global comparisons, again colored for within row comparison. Globally, our analysis indicates that the Michigan-based ANES comes closest to

Census marginal demographic distributions, and that our CCES sample is second. But CCES is more similar to the telephone-based NAES than to the face-to-face ANES.

To conclude that the RDD sampling methodology has generated the least representative sample (NAES) would be somewhat unfair, as both the ANES and the CCES provide sample weights matching to demographics and the NAES does not. However, the exercise does suggest that, except for the very high costs of face-to-face interviewing, the multi-stage area probability sampling design used by the ANES would be the preferred sampling method. The exercises further indicate that random digit dial telephone surveys are no panacea, and that non-random Internet panels such as the CCES should not be dismissed out of hand.

Analysis of political variables

In this section we compare marginal sample distributions of party identification, ideological self-identification, and political information. We selected these variables because we believe they may mediate micro-level persuasion effects. We also compare relationships between variables in different samples. Specifically, we examine the level of ideological constraint between party identification and ideology, as mediated by political information (Converse 1964; McGuire 1969; Zaller 1992).

The CCES survey carried three open-ended information questions of the kind the ANES has long used to assess political awareness. These questions ask what job or office Dick Cheney, John Roberts, and Dennis Hastert currently hold. The Hastert item produced much higher levels of awareness than did the parallel item in the 2004 ANES sample, but we believe this result is better explained by the extensive news coverage of Hastert over the Mark Foley scandal than by sample bias. Responses to the other two items are fairly close to results from the ANES. In the 2004 ANES, 85 percent of respondents correctly identified the job of Dick Cheney; in our sample, the number was 93 percent.[3] This difference, although perhaps substantively

[3] Unless otherwise noted, all statistics presented are derived using each survey's sample weights.

uninteresting, is statistically significant at the p < .01 for a two-tailed test. Twenty-eight percent of ANES versus 27 percent of our respondents correctly identified the job of William Rehnquist or Roberts. We should add, however, that the Roberts-Rehnquist comparison is difficult to evaluate. Roberts took over as Chief Justice shortly before the CCES survey, whereas in 2004 Rehnquist had been Chief Justice since 1986. When Rehnquist was new on the job in 1986, only 18 percent of ANES respondents identified his job.[4]

Figure 1 plots the distribution of an additive scale of the two more comparable items in the CCES 2006 and the ANES 2004. There are more than twice as many individuals who answered both the Cheney and Rehnquist-Roberts items incorrectly in the ANES sample.

Figure 2 plots the distribution of ideology in the CCES and the ANES. Unfortunately the two items are not directly comparable, for two reasons. First, the ANES uses a seven-point scale and the CCES uses five. Without knowledge of how the ANES categories should collapse to five, strong conclusions cannot be drawn from comparison of the distributions. The second difference is that the ANES invites responses of "haven't thought much about that," whereas the CCES does not. As a result, about 23 percent of ANES respondents are missing on this item, compared to 3 percent on CCES. Given this question difference, one might expect the ANES respondents who do report an ideological self-identification to report more extreme identification than the larger number of CCES respondents who offer self-reports (since presumably some of these people are making a best guess at their ideology). As we shall see, however, the CCES respondents appear to be more constrained in terms of ideology, suggesting that their self-identified ideology is doing more work in terms of helping them make sense of politics than the ideology of NES respondents.

In Figure 3 we present the distribution of partisanship in the two surveys, this time on the same scale and with the same question. The CCES sample holds more "strong" partisans.

[4] No post-stratification sampling weights are provided with the 1986 ANES.

Strong Democrats and strong Republicans are 18.5 and 20 percent in the CCES, 16 and 17 percent in the ANES. At the same time, CEES has about one percent point more independents. Although the differences are not substantively large, the distributions are statistically different at $p = 0.04$. If the party identification scales in the two surveys are converted to three-point measures of strength of party identification, the CCES respondents are more likely to be strong partisans than the ANES respondents, $p = 0.003$.

We now turn to comparisons with other nationally representative surveys. It is possible that the small differences between the 2004 ANES distributions and the 2006 CCES distributions could be due to temporal changes, not to failures in the representativeness of the CCES survey. We have compiled four Pew Center telephone surveys from that intervening period plus the 2004 and 2006 election exit polls to note changes in distributions.

In Figure 4 we compare the distribution of ideology across all of the surveys. There are both temporal and question-wording variations in these data, so we collapse all scales to three points. The different surveys are plotted across the x-axis, and above each survey are the liberal, moderate and conservative sample proportions. The four Pew surveys are connected by solid lines as they have question wording comparability.

Figure 4 shows that the collapsed CCES distribution is consistent with the distribution of voters (media exit polls) and the distribution of citizens (Pew research center telephone surveys). In fact, the CCES maintains the rank order of ideology relative to the 2006 exit poll while the ANES does not (with the 2004 exit poll). The trend in the Pew data does not indicate change in ideology across the two-year bridge, suggesting a comparison between ANES and CCES on this variable is reasonable. A χ^2 test for independence between the 3-point ANES and CCES ideology distributions yields a statistically significant difference, $p = 0.003$, though in this case, it is the ANES that is more polarized.

In Figure 5 we present these same distributions for party identification, again collapsed to a three-point scale.[5] As before, the CCES has a few more independents than Pew, and for the collapsed scale, appears consistent with the ANES. The χ^2 test does not indicate a statistically significant difference of distributions CCES to ANES, $p = 0.246$.

As a second cut at understanding partisan commitment of the CCES sample, we examined expected relationships between political variables. Ideology and partisanship are "constrained" at the individual level, but to varying degrees. We expect that the least politically aware respondents will have a somewhat noisy mapping of ideology into partisanship. Among the most aware respondents, we expect to find ideology mapping well to partisanship. Figure 6 presents the relationship of ideology and partisanship across three levels of political information in the ANES, as created by the sum of correct identification of Cheney and Rehnquist jobs. The three panels are respondents with 0, 1 and 2 correct responses to the two questions. The x-axes plot self-placed ideology and the y-axes self-placed partisanship. Points are medians and bars extend to the 25th and 75th percentile responses.

The figure demonstrates constraint across all information categories, but with fewer departures from monotonicity of medians at the highest level of political awareness. Strong constraint would show a monotonically increasing relationship between ideology and partisanship. Departures exist in all three categories, but most so in the 0-correct response category. It is the individuals who do not map well between these constructs that we hope to find in the CCES sample.

Figure 7 presents the same graphics for our CCES Internet sample. Because of the 7-point versus 5-point ideology scales, the ANES and CCES graphics are not perfectly comparable. Even so, looking across the panels of Figure 7 reveals strong constraint in the

[5] The exit poll data are suppressed due to large inconsistencies in marginals, likely due to question wording.

CCES sample, even at the lowest level of political awareness. In each frame the medians are monotonically increasing. We are wary of drawing strong conclusions from these graphics. It is possible that the information scales we use to differentiate political awareness, constructed with the Cheney and Rehnquist/Roberts items, are inadequate. The three information groups are not equally sized and the scale may poorly distinguish respondents. Still, the high level of constraint in the CCES, particular among the least informed, suggests that the Internet sample may be more ideologically engaged than the NES face-to-face sample.

To improve upon the simple, 2-item additive scale of the previous test, we ran Rasch item-response models on information items in the ANES and CCES surveys.[6] Item-response models are used extensively in psychometrics and standardized testing contexts such as the SAT and GRE, to estimated latent individual traits such as "intelligence" from responses to a set of survey questions. The models are especially effective because they utilize the discrete nature of survey questions and do not assume normality, which latent trait-estimating methods such as factor analysis and principal components analysis assume. We use an item-response model to estimate the latent respondent trait "political awareness" from responses to factual political information questions in the two surveys.

To test for constraint, we regressed partisanship on the information scale, 3-point ideology, and the information scale interacted with ideology. We present results of this parametric analysis in Table 1 for all respondents and for voters only.

In Figure 8 we plot predicted values at different levels of information and ideology using the parameters from the all-respondent models (Table 1, columns 1 and 3). The figure provides further evidence of greater constraint among the less informed in the CCES sample. While knowing the 3-point ideology of low political awareness respondents in the ANES (left panel) provides little information about their partisanship, the predicted partisanship of the lowest

[6] This includes 15 items for the CCES and 4 items for the ANES.

information CCES liberal is more than 2 points lower on the 7-point party identification scale than for the lowest information CCES conservative (right panel). This difference runs against the grain of another artifact: the CCES information scale has more items than the ANES scale and ought therefore to do a slightly better job of discriminating respondents who are truly low on political information and, therefore, ideologically unconstrained.

Conclusions

Compared to the 2004 ANES sample, we find that the CCES sample may be better informed: half as many people were unable to identify the jobs of Roberts and Cheney in the CCES as in the NES (7 percent v. 15 percent respectively). And, results of the item response model suggest that the lower levels of political awareness are missing in the CCES in terms of the constraint between ideology and party identification (Figure 8). Further, we find that the CCES has more strong partisans as measured on the traditional 7-point scale, but when we collapse these measure to their 3-point form, the distributions are not distinguishable from one another. Finally, we provide evidence that CCES respondents across all levels of political awareness exhibit constraint (as measured by the relationship of ideology to party identification) as compared to the NES respondents who only show this constraint at higher levels of political awareness – this again (Figures 6 and 7) suggests to us that respondents at lower levels of political awareness are missing in the CCES.

Across these three comparisons, the differences are not large, but neither do they seem entirely ignorable. Standard theories of persuasion suggest that a sample having these biases would exhibit less susceptibility to persuasion at a given level of exposure to persuasive messages. Despite the relatively close match between CCES and NES in terms of marginals on political variables, the differences in the way known mechanisms perform at the low levels of political awareness in the CCES may make it difficult to find advertising effects as we turn our attention to studying more than 60 Congressional advertising campaigns in the Midwest.

In a world without resource constraints, the results of our analysis would recommend against use of opt-in Internet samples for studies of political persuasion. In the real world, however, the recommendation is not so clear. Although we have not fully analyzed the representativeness of the NAES telephone sample, our partial analysis, based on demographic comparisons with census data, suggests that the bias of RDD sampling methods may be comparable to the bias of a Polimetrix generated (using sample matching) Internet sample. And telephone surveys, though less expensive than face-to-face, are much more expensive than opt-in Internet samples. Thus, our surprising but -- we stress -- preliminary conclusion is that Internet samples may be preferable to telephone surveys for some kinds of studies.

Scholars are naturally wary of relying on biased samples -- whether from Internet or telephone interviews. But data bias is not the only source of error in a study, and we believe this is worth remembering as we move into a new era of survey research. The random error due to the small sample sizes of expensive surveys like the ANES is a major concern. Larry Bartels (1985) points out that a biased but efficient estimator can have a smaller Mean Squared Error than an unbiased but inefficient one. By the same logic, a mildly biased but large Internet survey can produce more reliable estimates than an unbiased but small survey. Precision, even with biased estimates, may be preferable.

But, this said, we add two important caveats. We are not yet sure the bias of Internet samples is mild; our preliminary results are encouraging but by no means definitive. Second, scholars who use Internet (or telephone) surveys should find ways to build acknowledgement of the bias into their analysis. When researchers get a null result from a small sample, they are typically not tempted to publish it as a firm finding. When researchers obtains a null result from a very large but biased Internet sample, they should show the same caution unless they have somehow built the effect of the likely bias into their analysis.

A reason for optimism about the potential for Internet surveys is that Polimetrix did not include political information or engagement as one of the variables in their sample matching algorithm, nor did they stratify or weight on political awareness or turnout history. All indications are that soon they will be able to include these dimensions in their sampling and weighting. Thus, as long as low information people exist in their large Polling Point panel, smaller samples generated using sample matching and accounting for political awareness may look even more like the ANES in terms of ideological constraint across information categories.

All survey methodologies have weaknesses, some known and others unknown. In a recent article, Malhotra and Krosnick (2007) argue that face to face interviewing produces *less* social desirability bias than phone surveys because the respondent develops a rapport with the interviewer who is in their home. The opposite argument seems just as likely – because this person is in front of you in your home, you are *more* likely to give an answer that does not make anyone in the room uncomfortable. Given that there are no "true" targets against which to assess the political marginal distributions about which we care, the decision to use any survey methodology must be met with caution and full disclosure about the strengths and weaknesses of the method – and how it might affect the results. Polimetrix's sample matching technique seems to produce Internet samples that look more similar to existing RDD phone surveys than to multi-stage probability sample face-to-face surveys. The question remains, however, whether the ignorability assumption holds such that the people who take surveys online behave the same way as those who do not – or more realistically in terms of measurement – as those who take phone or in-person surveys.

References

Bartels, Larry B. 1985. "Alternative Misspecifications in Simultaneous-Equation Models." *Political Methodology* 11 (3-4):181-199.

Converse, Philip. 1964. "Nature of Belief Systems in Mass Publics," in David Apter (ed) *Ideology and Discontent*. Free Press.

Curtin, Richard, Stanley Presser, & Eleanor Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69 (1): 87-98.

Freedman, Paul, Michael Franz, and Kenneth Goldstein. 2004. "Campaign Advertising and Democratic Citizenship." *American Journal of Political Science* 48:723-41.

Goldstein, Kenneth. 2006. *Cooperative Congressional Election Study*. Wisconsin Team Module. Palo Alto, CA: Polimetrix, Inc.

Hill, Seth J. 2006. "Campaign Events and Political Mobilization." Working Paper, UCLA.

Hillygus, D. Sunshine, & Simon Jackman. 2003. "Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy." *American Journal of Political Science* 47 (October): 583–596.

Hillygus, D. Sunshine. 2005. "Campaign Effects and the Dynamics of Turnout Intention in Election 2000." *Journal of Politics* 66: 50-68.

Jackman, Simon. 2006. *Cooperative Congressional Election Study*. Stanford Team Module. Palo Alto, CA: Polimetrix, Inc.

Johnston, Richard, Michael G. Hagen, and Kathleen Hall Jamieson. 2004. *The 2000 Presidential Election and the Foundations of Party Politics*. New York: Cambridge University Press.

Malhotra, Neil and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis Advanced Access* February.

McGuire, William. 1969. "The Nature of Attitudes and Attitude Change." *The Handbook of Social Psychology* 3: 136-314.

Vavreck, Lynn. 2006a. *Cooperative Congressional Election Study*. Polimetrix Module. Palo Alto, CA: Polimetrix, Inc.

Vavreck, Lynn. 2006b. *Cooperative Congressional Election Study*. UCLA Team Module. Palo Alto, CA: Polimetrix, Inc.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion.* New York: Cambridge University Press.

Zaller, John. 1996. "The Myth of Massive Media Impact Revisited." In *Political Persuasion and Attitude Change*, ed. Diana C. Mutz, Paul M. Sniderman, & Richard A. Brody. Ann Arbor: University of Michigan Press.

Table 1. Comparison of Sample Demographics to the Census.

	Census 2004	NAES 2004	NES 2004	CCES 2006
Sample Weights Range	-	-	.36 - 3.02	.33 - 2.98
Education	Census	Annenberg	ANES	CCES
Less than HS	**16.1**	7.3	14.4	3.9
HS	**29.5**	25.6	31.6	42.3
Some College Through 4 Year Degree	**44.5**	51.7	44.1	45.1
Advanced Degree	**9.9**	14.3	9.9	8.7
Age	Census	Annenberg	ANES	CCES
18-24	**12.3**	7.7	11.5	8.6
25-34	**18.1**	16.1	17.2	17
35-44	**20.5**	20.3	19.4	25
45-54	**19.3**	21.6	19.2	30.8
55-64	**13.6**	16.3	15.8	10.6
65+	**16.1**	17.9	16.9	8.1
Race	Census	Annenberg	ANES	CCES
White (Non-Hispanic)	**67.3**	78.2	70.9	71.1
Black	**12.8**	8	14.8	10.9
Hispanic	**14.1**	5	6.3	12.2
Other	**5.8**	8.8	8	5.8
Gender	Census	Annenberg	ANES	CCES
Male	**48.9**	44.7	49.5	48.2
Female	**51.1**	55.3	50.5	51.8

Cell entries are percentages within the given category, weighted for the ANES and the CCES. Annenberg values are calculated from the sum of Cross-Sections data file across 2004. Census distribution from the 2004 American Community Survey.

Table 2. Deviation Statistics of Sample Demographics Relative to the Census.

	NAES 2004	NES 2004	CCES 2006
Education Comparisons			
Mean Absolute Deviation	6.1	1.1	6.7
Mean Deviation/Actual	0.32	0.05	0.33
X^2 Test P-Value	0.004	0.952	0.000
Age Comparisons			
Mean Absolute Deviation	2.3	1.0	5.3
Mean Deviation/Actual	0.16	0.07	0.32
X^2 Test P-Value	0.56	0.99	0.01
Race Comparisons			
Mean Absolute Deviation	7.0	3.9	1.9
Mean Deviation/Actual	0.42	0.29	0.08
X^2 Test P-Value	0.000	0.013	0.842
Gender Comparisons			
Mean Absolute Deviation	4.2	0.6	0.7
Mean Deviation/Actual	0.08	0.01	0.01
X^2 Test P-Value	0.40	0.90	0.89
Global Comparisons			
Mean of Absolute Deviations	4.6	1.7	4.2
Median of Absolute Deviations	4.2	1.0	2.4
Standard Deviation of Absolute Deviations	2.9	1.9	4.4
Mean of Deviation/Actual Ratios	0.26	0.11	0.22

P-values from the χ^2 statistic comparing the Census demographic distribution as truth to the sample demographic distribution.

Table 3: Party Identification, Ideology, and Information

Weighted Least Squares regression parameter estimates (with standard errors in parentheses) for dependent variable 7-point party identification by survey sample.

	ANES 2004	ANES 2004 (Voters Only)	CCES 2006	CCES 2006 (Voters Only)
Information Scale	-1.06*** (0.19)	-0.92*** (0.24)	-0.57*** (0.12)	-0.43*** (0.14)
3-pt ideology	1.29*** (0.07)	1.33*** (0.08)	2.06*** (0.05)	2.07*** (0.05)
Ideology*Info	0.66*** (0.08)	0.61*** (0.11)	0.28*** (0.05)	0.24*** (0.07)
Intercept	*1.17**** (0.16)	*1.02**** (0.20)	*-0.37**** (0.10)	*-0.39**** (0.12)
R^2	.32	.32	.52	.53
N	1012	713	1885	1426

*p < .10 ** p < .05 *** p< .01

Figure 1:

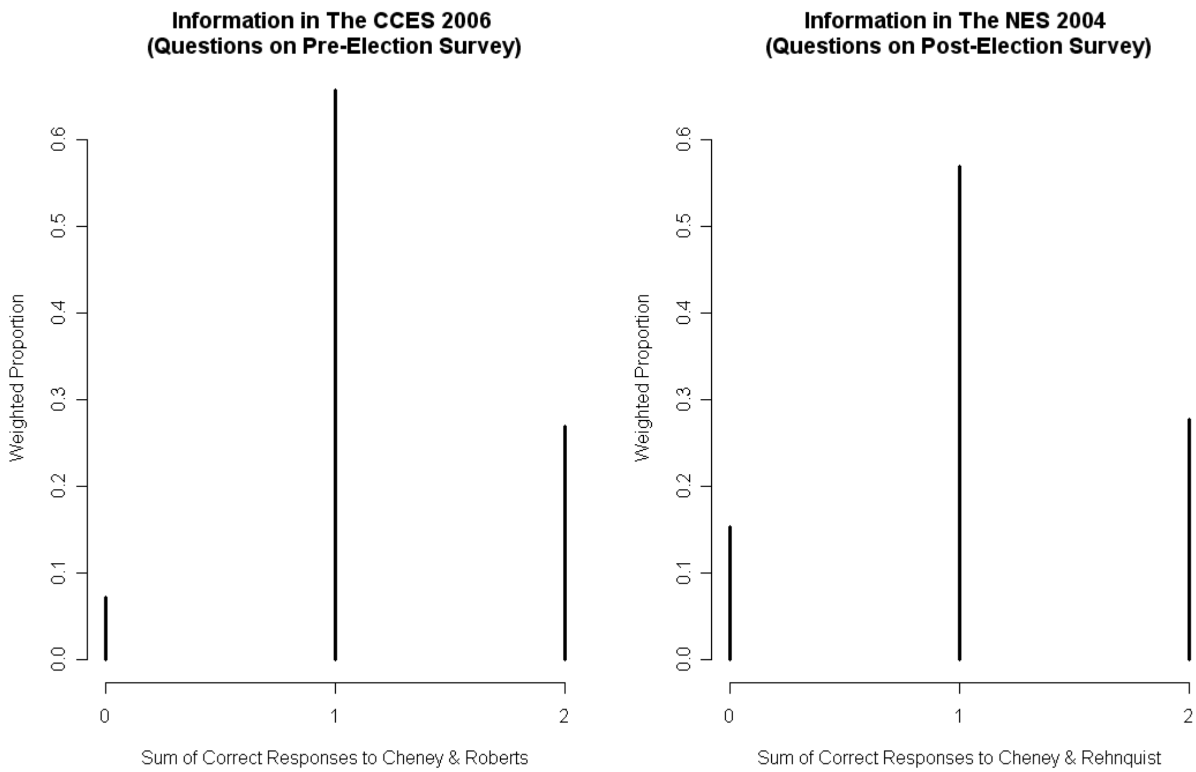


Figure 2:

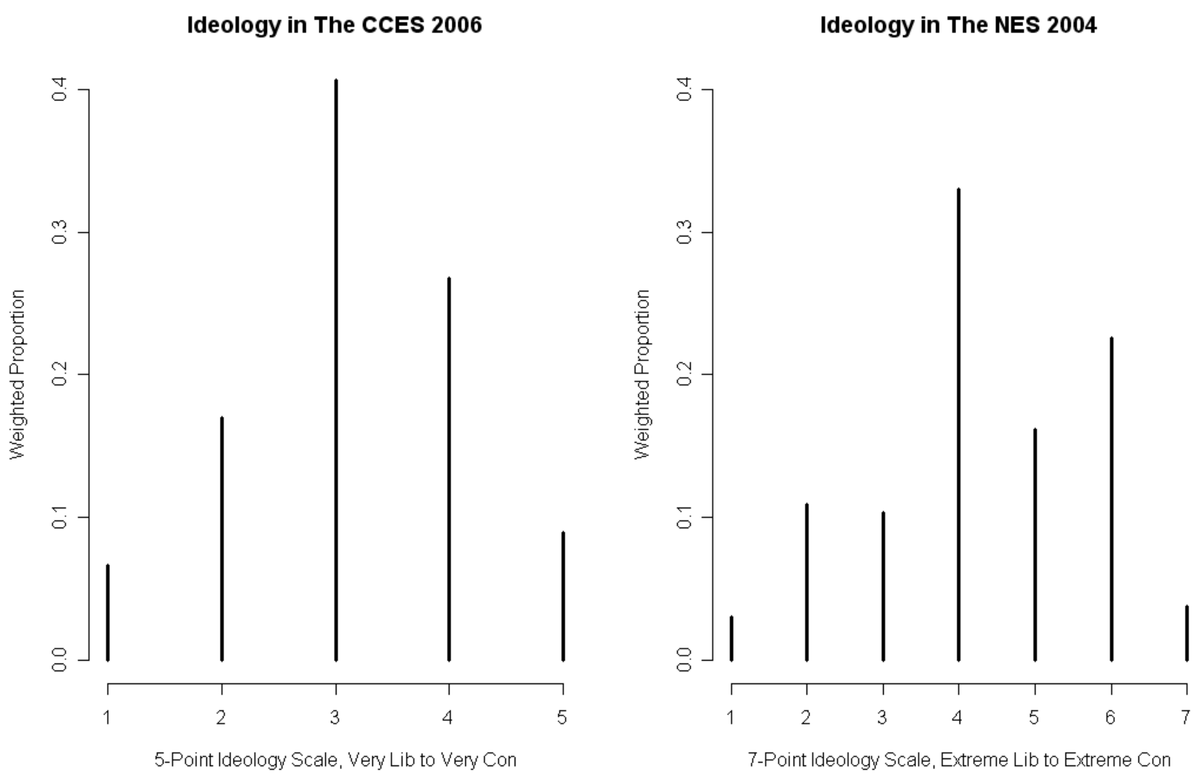


Figure 3:

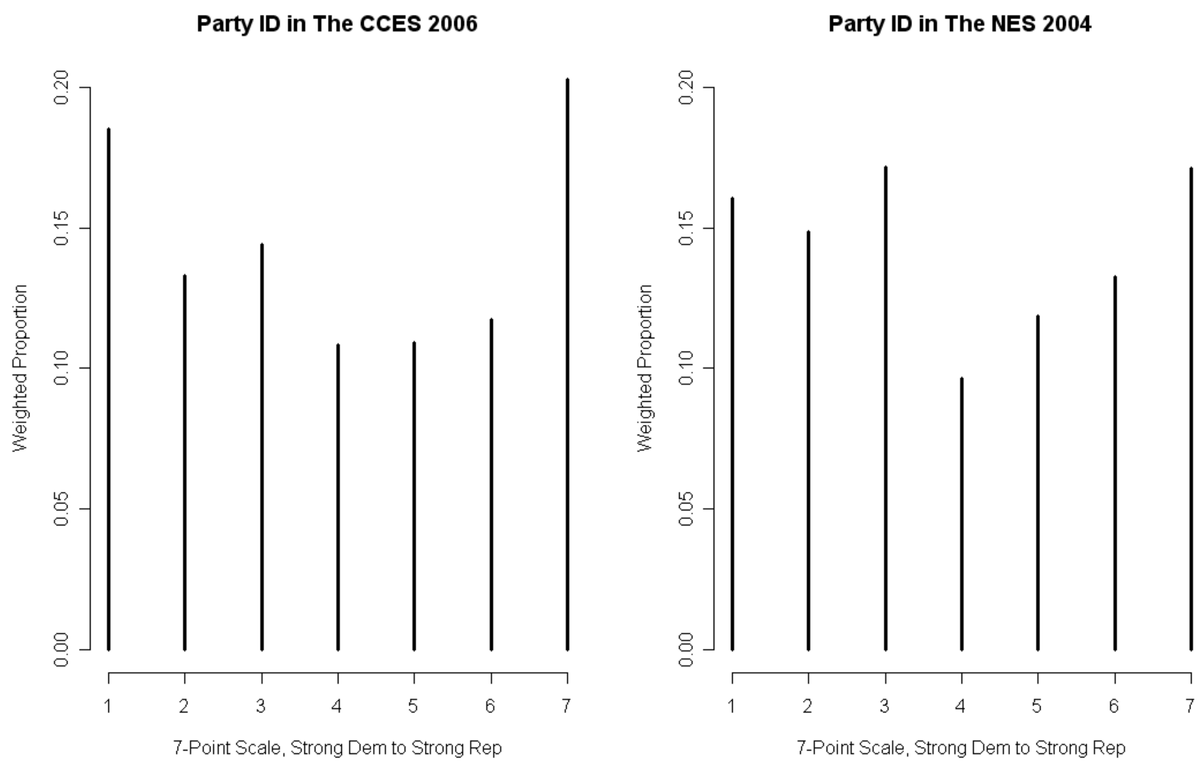


Figure 4:

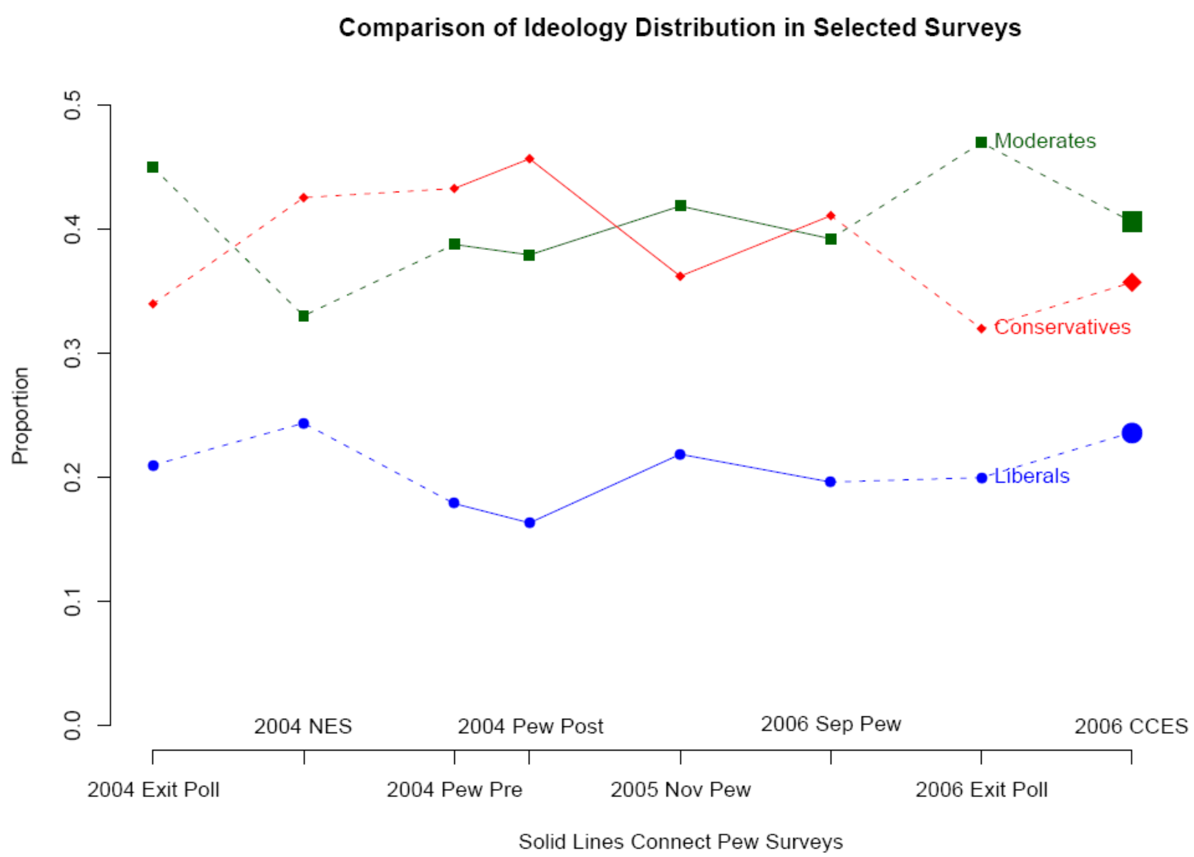


Figure 5:

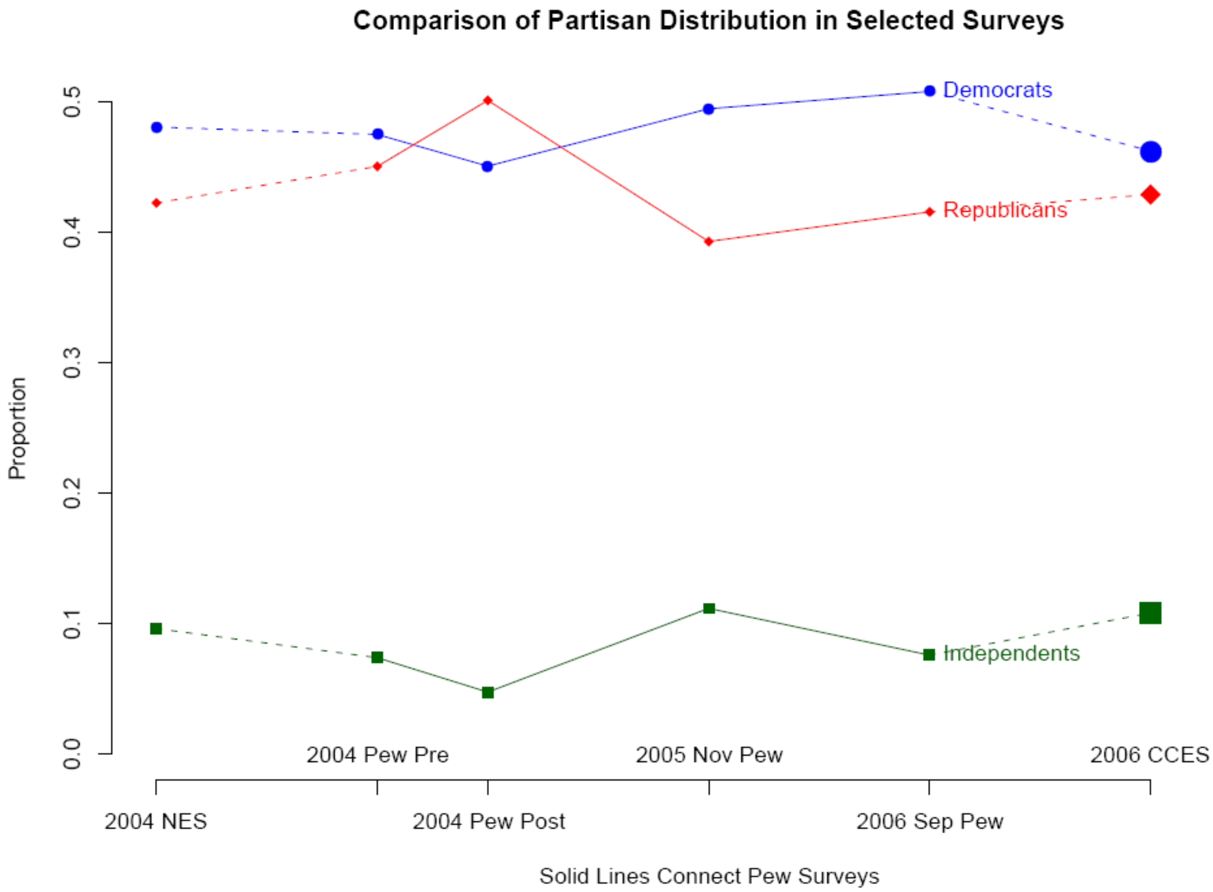

Comparison of Partisan Distribution in Selected Surveys
Democrats
Republicans
Independents
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
2004 Pew Pre
2005 Nov Pew
2006 CCES
2004 NES
2004 Pew Post
2006 Sep Pew
Solid Lines Connect Pew Surveys

Figure 6:

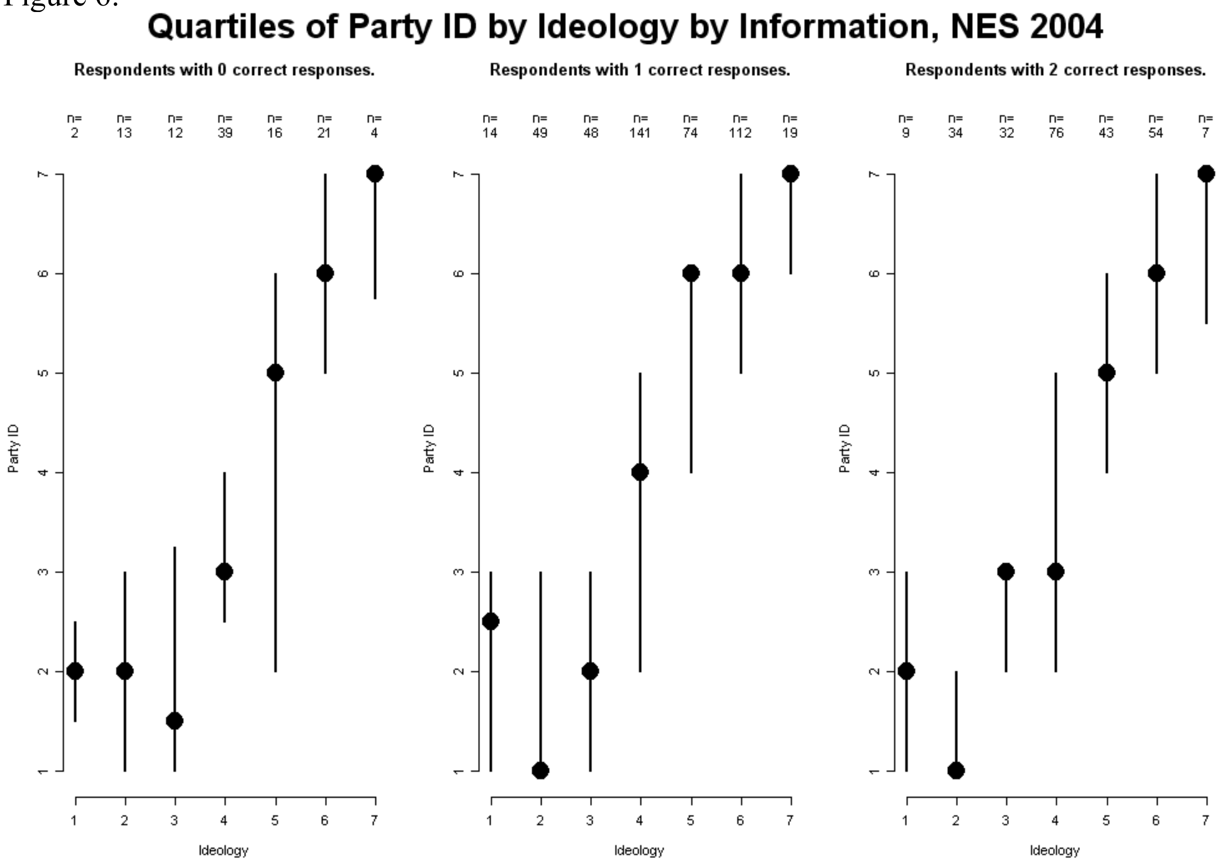


Figure 7:

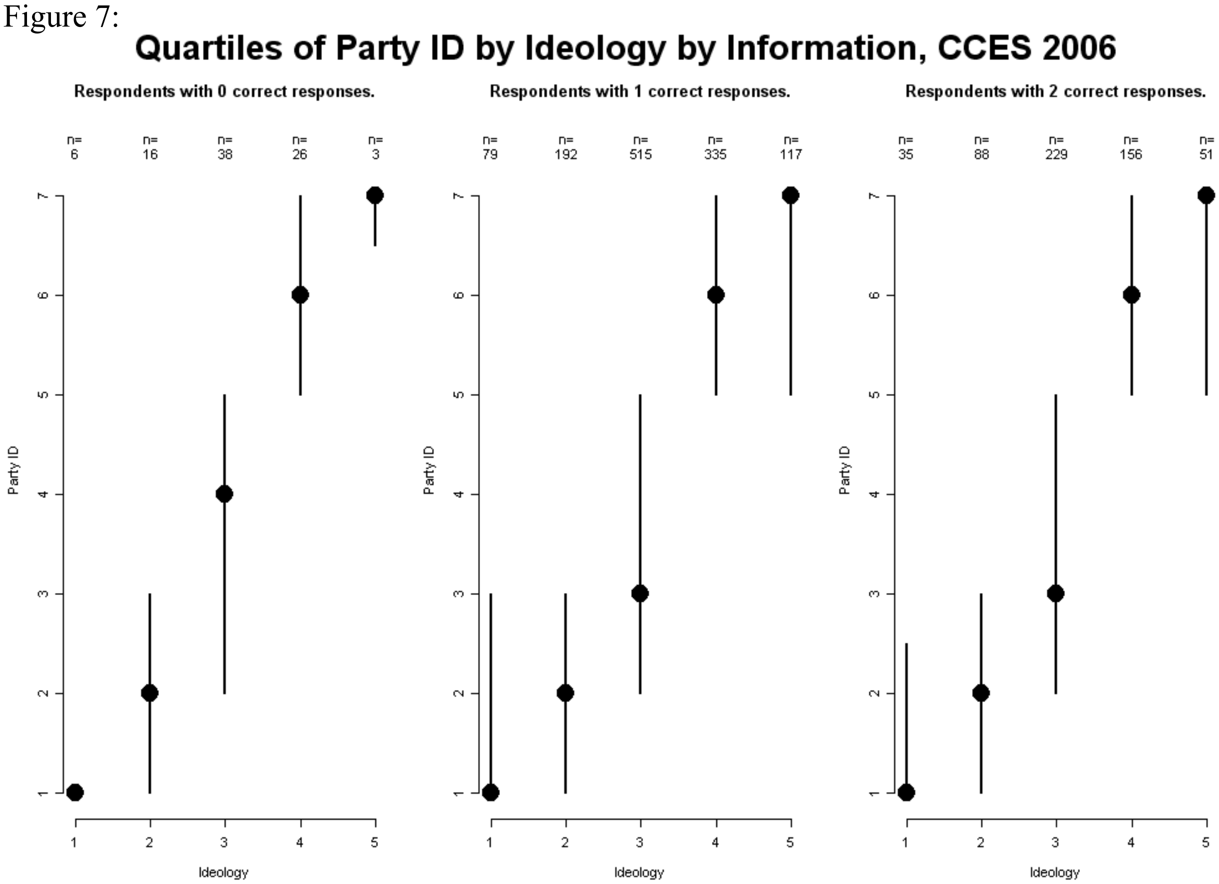
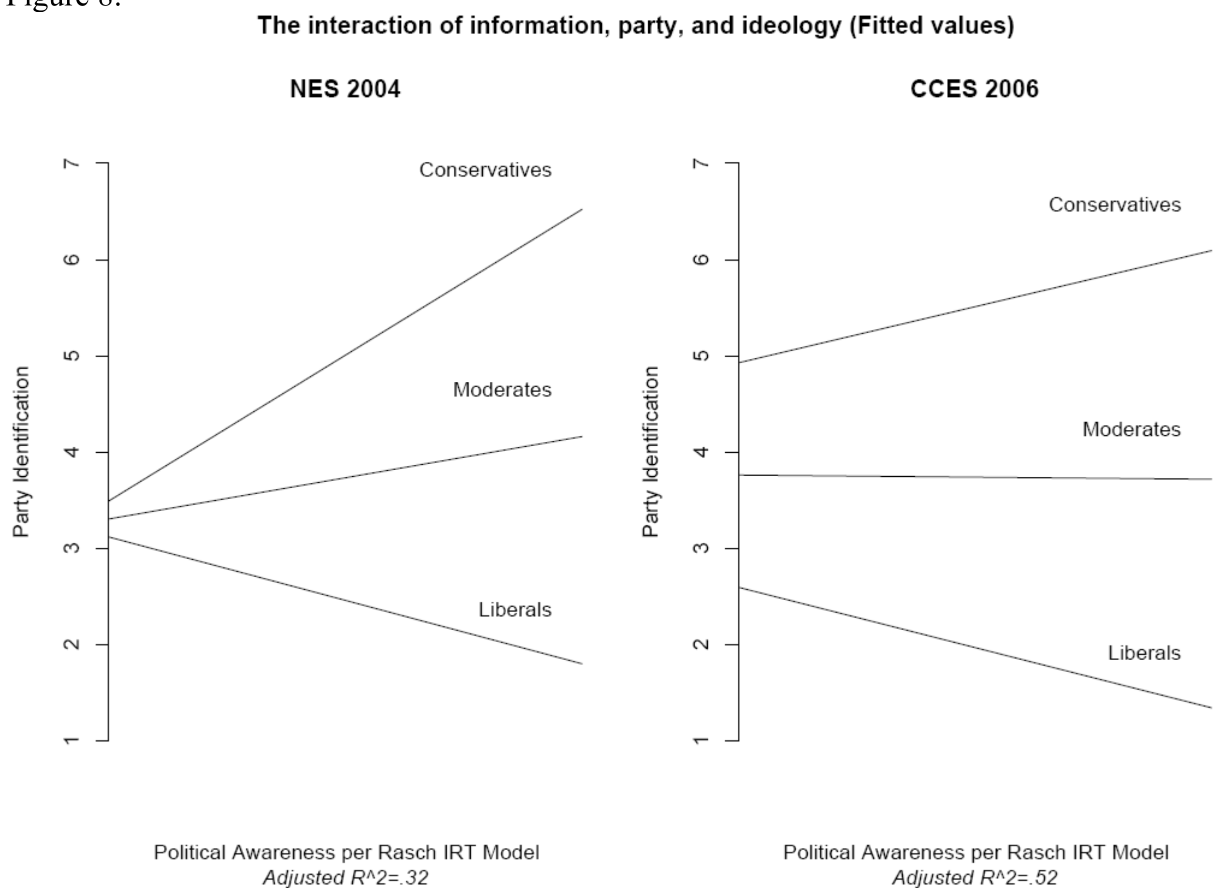


Figure 8:


The interaction of information, party, and ideology (Fitted values)
NES 2004
CCES 2006
Party Identification
1
2
3
4
5
6
7
Conservatives
Moderates
Liberals
Political Awareness per Rasch IRT Model
Adjusted R^2=.32
Adjusted R^2=.52