

Experimental Tests of the Question-Answering Model of the Mass Survey Response*

by

John Zaller
University of California, Los Angeles

2-22-90

* This paper is a technical report to the National Election Studies Board of Overseers on results obtained from the 1989 Pilot study. The paper is based on data collected by the National Election Studies under a grant from the National Science Foundation. I alone bear responsibility for the analysis and conclusions reached in the paper.

Abstract

This paper reports on experimental tests of two hypotheses derived from the new question-answering approach to understanding mass political attitudes. The hypotheses are, first, that questions which frame issues differently will produce shifts in the correlates of attitudes on these issues, and second, that questions which provide issue frames will enhance response stability. The first hypothesis was supported in two of four experimental tests; the second was also modestly, though unevenly, supported. These results suggest the possibility of both practical improvements in the measurement of political attitudes, and improved understanding of the nature of attitudes.

Much recent attitude research has asserted that individuals do not possess pre-formed opinions on every issue on which pollsters happen to inquire. Instead, persons are assumed to construct their attitude reports on-the-fly as they move through the questionnaire. This memo reports on tests of two experimental hypotheses derived from this new "question-answering approach" to the mass survey response.

The first hypothesis, following work by Kinder and Sanders (1990), is that changes in question frame can produce changes in the patterns of support for particular issues, even when the frame manipulation leaves the core issue and response options unchanged. This hypothesis was supported in two of four tests. In one of the supporting cases, religious people were more likely than other persons to support increased funding for the Contra rebels when the question framed the issue in terms of opposition to communism; when the issue was framed differently, religious orientation had no effect on attitudes toward Contra aid.

The second hypothesis is that questions which contain a supporting argument for each option, as compared to items which consist of spare response alternatives and nothing else, will anchor people's responses to the particular arguments used, thereby increasing over-time response stability. This hypothesis obtained modest support. The gain in response stability across five experimental comparisons averaged about 14 percent, an overall effect which was just statistically significant. But in only one of five cases was the increase in response stability statistically significant at the level of an individual item, and in one case there was a slight reversal.

In addition, in all five cases in which supporting arguments were added to short-form questions, rates of no opinion fell, and in the two cases in which the no opinion rates were initially above 15 percent, the reductions were substantial. Thus, argument-anchored items hold out the practical promise of modestly increasing response stability on some items while also reducing no opinion rates.

The results of these experiments have both practical and theoretical value. On the practical side, the modest gains in stability, along with the lower rates of "don't know" responses that were obtained for all five comparisons of framed and unframed questions, suggest that, in at least certain cases, framed questions may make better overall measures of political attitudes. On the theoretical side, the results provide additional evidence that many people do, as the question-answering model asserts, formulate their attitude reports in response to the particular stimuli confronting them -- even if, as the uneven quality of the results also indicate, we cannot yet predict how exactly they go about doing so.

Design of Study

As indicated, the study was designed to test two hypotheses -- that questions which frame issues differently will appeal to different types of people, and that questions which provide issue frames will enhance response stability. The items necessary to test these hypotheses partially overlap. Two (of the four) forms of the survey carry differently framed items on four different issues. Comparison of the correlates of these differently framed item pairs provide the test of the first hypothesis. A third form of the study carries unframed or "stripped" versions¹ of five items, each of which has a counterpart framed version on another form of the questionnaire. (Four of the framed counterparts are the items used to test the first hypothesis, while the fifth is new.) Comparisons of overtime stability rates for the five framed and unframed items are the test of the second hypothesis. (A fourth form of the survey administers one framed item to respondents on the first wave of the survey and the other framed item on the second wave; however data from this form are not used in the paper.)

Experiment 1: Changing Patterns of Correlation

¹ The apt distinction between framed and stripped items owes to Donald Kinder.

In the course of developing new items to measure racial attitudes, Kinder and Sanders (1990) found that patterns of support for affirmative action depended on the way the question was framed. When the argument against affirmative action was that blacks would obtain rewards they had not earned, whites scoring high on "symbolic racism" were most opposed to it. But when the question was framed in terms of unfair disadvantage to whites, white persons perceiving themselves to be in competition with blacks were most likely to oppose affirmative action; attitudes on symbolic racism, meanwhile, ceased to have important effects.

These correlational shifts can be explained as follows: Individuals often have mixed feelings on political issues. Which of their feelings becomes the basis of their survey responses depends on the ideas the question itself has made salient. If the question makes salient an idea that is held most frequently by a certain type of person -- say, persons perceiving themselves to be in competition with blacks -- then perceived feelings of competition with blacks will become correlated with responses to the question. But if the question raises some other idea instead, feelings of competition will not be activated as determinants of answers to the question.

Thus, the shift in the pattern of support for affirmative action supports the central claim of the question-answering approach to understanding mass attitudes, which is that individuals construct their attitude reports as they encounter each survey item, and that they do so on the basis of the ideas that are most immediately salient to them (Tourangeau and Rasinski, 1987; Zaller and Feldman, 1988).

But although theoretically provocative, the Kinder and Sanders result was obtained by accident and has not yet been replicated. To test both its resilience and its generality, I developed four pairs of items which might be expected to appeal differently to different types of persons, thereby producing a Kinder-Sanders type of shift in the correlates of policy support. These items, each of which involved both a pro and a con argument, were

- Funding for the B2 bomber. In one item, the pro-B2 argument was that it was needed to protect the U.S. from communism. In the other item, the anti-B2 argument was that it would fuel an "immoral arms race." Because I thought religious persons would be responsive to both the communism and immorality frames, I expected them to support the B2 more strongly on the first item than on the second.

- The death penalty. In the first item, the pro-death penalty argument was that murder is a crime which "deserves" death. In the second item, the anti-death penalty argument was that the taking of another's life is always immoral. Because I thought religious persons would be responsive to both the Old Testament justice and immorality frames, I expected them to support the death penalty more strongly on the first item than on the second.

- Support for the Contra rebels in Nicaragua. One item stressed "freedom fighters" who were trying to prevent the spread of communism, while the other stressed support for a group that was "fighting to promote democracy." I expected religion to be more strongly correlated with the item which mentioned communism.

- Oil drilling in the Alaskan wilderness. In one item, the pro-drilling argument was the danger to "working people" of "higher gasoline prices and further layoffs in American factories," while in the second item, the pro-drilling argument was the need to "avoid becoming dependent on foreign oil." I expected a person's identification as working class or middle class to be more strongly correlated with responses to the first item than the second.

It should be emphasized that all response options were identical in both forms of each question. All that changed were the reasons given for taking or rejecting each option -- that is, the way the issue was framed. Exact question wordings are shown in the Appendix.

Measurement of one of the key variables, religious sentiment, proved difficult. The most obvious measure, church attendance, produced an effect which ran in the expected

direction, but which did not approach statistical significance. The problem, I suspected, is that some people attend liberal or non-traditional churches, in which case the expected sensitivity to anti-communism would not be present.

Fortunately, the 1988 study contained questions which could be used to measure the extent to which a person was not only religious, but religiously conservative. For Protestants, there was a question asking how well a person's religious views could be described as fundamentalist; for Catholics, the comparable question asked about "traditional" religious views. These questions were supplemented by a third item concerning very frequent Church attendance (more than once a week). Persons who professed no religious beliefs were placed at the lowest category of the religion measure, indicating no religious commitment. Persons who were religious but were neither Protestant nor Catholic were omitted since the nature of their religious views could not be assessed.²

Social class identification was measured from a simple self-report item administered to all respondents (v1208).

Religion and social class identification represent auxiliary attitudes which might or might affect responses to the given issue, depending on the framing of the question. For the B2 and Contra items, I expected a person's general attitude toward defense spending to be the primary determinant of policy attitudes; for the oil drilling item, I expected a feeling thermometer on environmentalists to play this role; and for the death penalty item, for lack of anything better, I relied upon a person's self-identification as a liberal or conservative as the primary determinant of policy preferences.

With these primary and auxiliary variables, I estimated the following model for each dependent item:

² The variables used in the religion measure are, respectively, V2538 and v2529, each a four point item; v1215, a two-point item; and v8205, where codes of 6 and 8 were counted as non-religious responses. The religion scale has a range from 1 to 6.

Table 1
Coefficients for Tests of
Shift in Attitude Correlates

	Contra aid	B2 bomber	Death penalty	Alaska oil drilling
Intercept	5.79	1.08	2.00	2.03
Primary Determinant*	0.48 (0.10)	0.44 (0.08)	0.51 (0.14)	0.04 (0.007)
Auxiliary Determinant**	-0.08 (0.17)	0.07 (0.14)	-0.13 (0.21)	-0.02 (0.19)
Form	-1.45 (0.78)	0.31 (0.64)	-0.43 (0.89)	-1.50 (0.69)
Form X Auxiliary	0.43 (0.24)	-0.13 (0.20)	0.16 (0.27)	0.51 (0.28)
r^2	.15	.17	.09	.16
N=	165	159	136	180

NOTE: Dependent variable in this analysis is sum of wave one and wave 2 policy variables (Contra aid, B2 Bomber, etc.). Persons having no opinion on policy item in either wave are treated as missing data. Standard errors appear in parentheses.

* The primary determinant of the four dependent variables are, respectively, defense spending attitudes (1-7), defense spending attitudes (1-7), ideological self-designation (1-7), and feelings toward environmentalists (1-7).

** For the Contra aid, B2 bomber, and death penalty items, the auxiliary determinants are religious orientation; for the oil drilling item, it is social class identification. See text for further discussion.

$$\text{Item} = b_0 + b_1 * \text{Primary} + b_2 * \text{Form} + b_3 * \text{Aux.} + b_4 * \text{Form} * \text{Aux.}$$

In the case of the item on Contra aid, for example, the primary variable was the person's defense spending attitude and the auxiliary item was the person's religious orientation.

The results of estimating the model are shown in Table 1. All variables except the interaction variables have been scored in the liberal direction; the interaction variables have been scored so that the expected interactions run in the positive direction. The dependent item combines responses to both the first and second wave administrations of the dependent item.

INSERT TABLE 1 ABOUT HERE

As can be seen in the table, the key interaction terms achieve statistical significance in the expected direction for only two items, aid to the Contras and oil drilling in Alaska. The other two interactions are non-significant, with one running slightly in the expected direction and the other slightly reversing expectations.

According to these results, an "average working class" person is about .75 scale units more likely to favor oil drilling in Alaska than an "average middle class person," but only when the danger of high gas prices and layoffs has been mentioned in the question. Since the dependent item is a four-point item running from strongly support to strongly oppose, the magnitude of the effect is substantial. Persons describing their views as fundamentalist and also attending church frequently were about one scale unit more likely (on a five-point scale) to favor aid to the Contra rebels than were non-religious persons, but only when the Contra item mentioned opposition to communism.

These results would obviously be stronger if correlational shifts had appeared in all four tests rather than in only two of them. Yet nothing in the question-answering approach requires that correlational patterns change with every change of question frame. It could be that an attitude dimension is fully engaged in determining people's survey responses even in the absence of a question frame to cue it, so that a more explicit cue would have no additional effect on the correlational pattern. Or it could be

that a shift in question frame affects all respondents rather than particular types of respondents, in which case no correlational shift would occur. Or it could be that an hypothesized auxiliary variable is simply irrelevant to people's attitudes (as was apparently the case for the B-2 and death penalty items). In short, the expectation of a correlational shift depends on a key auxiliary condition -- the ability to find a frame that will affect a particular type of person more than other types -- that cannot easily be directly verified. As long as this is the case, it may be unrealistic to expect every attempt to produce correlational shifts to be successful.

Experiment 2: Increasing Item Reliability

In my proposal to the NES Board, I argued that one reason for the low reliability of attitude items is that typical survey questions make unrealistic demands of respondents. That is, survey questions provide too few cues about what issues are about, and too little time to search one's memory for pertinent thoughts and beliefs.

I therefore proposed to experiment with the effects of questions which contained a short argument in support of each response option. The greater length of these questions might afford people somewhat more time to ponder the issues before venturing an opinion, thereby enhancing overtime response stability. And the particular arguments would, I thought, anchor people's attitude reports to particular considerations, thereby further enhancing stability.

The test of these expectations involved comparison of overtime stability rates with framed items, as described in the previous section, with unframed or stripped items. Typical of the framed and unframed items used in these comparisons are the following:

Drug Testing -- Framed

Some people believe that illegal drugs like heroin and cocaine are the greatest problem facing America today. As one step in dealing with this problem, they favor giving all employers the right to test their workers for use of illegal drugs. Other people oppose such testing as an invasion of privacy. They say that as long as a worker is performing well on the

job, employers should not be able to require a drug test. Do you favor or oppose giving all employers the right to test their workers for possible use of illegal drugs?

Drug Testing -- Unframed

Do you favor or oppose giving all employers the right to test their workers for possible use of illegal drugs?

Note that the response options are identical for the framed and unframed items.

Comparisons were made across items measuring the four items previously described, plus the pair of items concerning the right of private employers to test workers for use of illegal drugs.

A comparison of overtime stability rates for the two types of items provide only modest support for my expectations. The measure of stability in these comparisons is the mean absolute change in scale position over the two waves of the survey. The average gain in stability for the five framed items was 14 percent, an overall difference that is just statistically significant on a one-tailed test ($p=.05$, one tailed). Full results at the item level are shown in Table 2. In only one case does a framed item produce an improvement that is statistically significant at the item level. That item, involving funding for the B2 bomber, produced a stability gain of 32 percent ($p < .02$, one tailed). One item, however, performed less well in its long form, namely the item on the death penalty. The problem may have been that responses to the death penalty item were, as Table 2 shows, extremely stable even in its short form, thus leaving scarcely any room for improvement.

INSERT TABLE 2 ABOUT HERE

Extra Words or Extra Ideas?

There are, as indicated, two possible explanations for why the framed items might have increased response stability -- that their length gave respondents more time to think, and that the arguments contained in the framed items anchored people's responses

Table 2

The Effects of Framed and Unframed
Items on Response Stability

Spending on B2 Bomber	Framed	.46*	p = .02
	n=	(180)	
Unframed	.68		
n=	(84)		
Oil drilling in Alaska	Framed	.51	p = .18
	n=	(200)	
Unframed	.60		
n=	(83)		
Aid to Contras	Framed	.37	p = .20
	n=	(216)	
Unframed	.43		
n=	(117)		
Death Penalty	Framed	.29	reversal
	n=	(228)	
Unframed	.22		
n=	(116)		
Drug testing of workers	Framed	.41	p = .14
	n=	(115)	
	Unframed	.52	
	n=	(122)	
	Verbose	.50	
	n=	(120)	

* Cell entries are mean absolute scale change across the two waves of the survey; all scales have four points, except Contra aid, which has five. P-values are one-tailed, as appropriate when testing a particular hypothesis.

to particular ideas. To test these competing possibilities, the Pilot study carried one long form item that was designed to be verbose but vacuous, that is, an item that would give respondents more time to think without also giving them additional arguments and ideas. This item, which was used in association with the earlier drug testing items, is:

Drug Testing -- Verbose

Here is a question about illegal drugs. As one step in dealing with this problem, some people think all employers should have the right to test their workers for possible use of illegal drugs. Others oppose this kind of testing. As you know there is a lot of debate in the press on this question of whether employers should have the right to test workers for illegal drugs. Which side of the debate is closer to your view? Do you favor or oppose giving all employers the right to test their workers for possible use of illegal drugs?

Comparison of the results obtained from these three items are contained in Table 2. Although the data are scarcely unequivocal, they provide no support for the notion that length per se enhances response stability.

Effect on Rates of No Opinion

The manipulation of question form did produce one consistent effect: Volunteered rates of "no opinion" were lower for the framed form items in all five comparisons with short form items, a result that could be expected, by a sign test, to occur by chance alone only three times in a hundred. However, for three of the items, no opinion rates were six percent or less on the short form items, so that large reductions in don't know rates could not easily be attained. But two of the short form items (defense spending and oil drilling) had no opinion rates of 23 and 27 percent, respectively; in the long form items, these rates fell to 16 and 12 percent, respectively, both highly statistically significant declines.

Thus one effect of using framed, as against unframed, items is to achieve higher levels of opinionation. The results in Table 2 suggest, in addition, that these higher response rates have been achieved without any loss in the quality of response.

(Although interviewers were instructed to immediately accept don't know responses, none of the items in these experiments explicitly offered respondents such a response. If, as is fairly standard practice, the experimental items had offered an explicit no opinion option, baseline no opinion rates would have been higher and the effect of using framed arguments might then have been greater.)

Individual Differences in Experimental Response

It is natural to wonder whether all types of respondents profited equally from use of the framed items. The answer, it turns out, is clearly no. Persons describing themselves as "extremely liberal" or "extremely conservative" did much less well using the framed items. Using the unframed items, these persons were virtually perfectly stable, changing by an average of just .06 scale units between waves of the survey. But using the longer or framed items, they changed an average of .36 units -- a highly statistically significant 600 percent increase in response instability. Although the small number of cases make the precise magnitude of this shift unclear, its existence is difficult to dispute.

Meanwhile, persons describing themselves as "strong liberals" or "strong conservatives" were 46 percent more stable using the framed items, a difference which is also highly statistically significant. Moderate liberals and conservatives, along with centrists, also appear to have made gains on the framed items -- though, owing to the reduction in cases, the gains are not statistically significant in either group alone. Finally, persons not knowing their ideological orientation may have become slightly less stable on the long form items. The overall pattern of these results is summarized in Figure 1.

INSERT FIGURE 1 ABOUT HERE

Unless replicated, it is probably a mistake to pay too much attention to all of the detail in Figure 1. The most probable interpretation of the results is that the framed form items are helpful (quite possibly equally helpful) to strong ideologues, moderate ideologues and centrists, and that they somehow interfere with the response processes of extreme liberals and conservatives. Thus, if one divides the sample into two groups -- extreme liberals and conservatives versus all others -- one finds a statistically significant loss in stability in the first group and a statistically significant gain in stability in the second group, even with two-tailed tests.

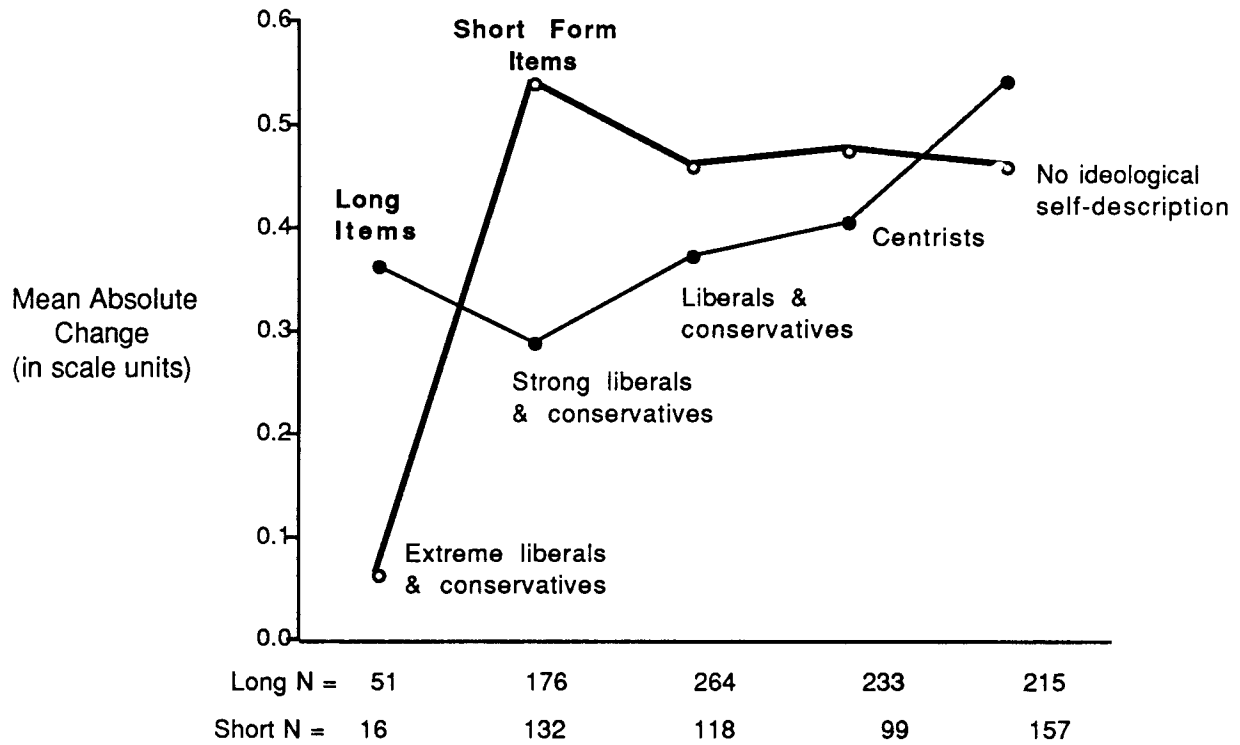
Despite its statistical significance, this empirical pattern might be suspect except that it replicates a similar finding from my earlier study of UCLA undergraduates (Zaller, 1989). In that study, I found that persons placing themselves at the extreme ends of an ideological self-identification scale preferred answering short form questions and also gave more stable answers when using these questions. The remainder of the sample expressed a preference for the longer form items and gave more stable answers when using them. This interaction was also statistically significant.

The existence of this interaction suggests that self-described extreme liberals and conservatives normally go about answering questions differently than other people do -- though it does not, unfortunately, indicate what this difference is. An obvious speculation is that persons willing to identify themselves as extremely ideological normally make heavy use of ideological cues in answering questions -- i.e., make "knee jerk" responses -- which, in turn, makes them more stable than persons who must try to formulate opinions on the merits of each issue. The additional argumentation of the framed items, on this view, forces even extreme ideologues to decide issues on the basis of substance, thereby introducing normal amounts of response unreliability.

Conclusion

Figure 1

The Effect of Long vs. Short Items on Different Ideological Types



Key

Short Form Items ———
Long Form Items ———

Practical Implications

Results from the second set of experiments make it doubtful that more verbally ample questions can dramatically reduce item unreliability in typical public opinion surveys. Krosnick's (1990) results, which show that the addition of very simple labeling to 7-point scales -- i.e., terms such as "very" and "somewhat" -- also fails to improve item reliability, reinforce this doubt.

Nonetheless, I believe it is too soon to abandon this line of investigation. The results reported in this paper and the results of my study of UCLA undergraduates both indicated that moderate gains are possible at least in some cases. For example, in both studies, stability gains of 40 percent or more were made from using long form questions about defense spending. It does not seem far-fetched to believe that, with further, patient developmental work, one can identify types of issues for which important gains in measurement quality can be obtained from using framed items.

The fact that the addition of frames can affect both marginal distributions (note the form effect on oil drilling and Contra aid in Table 2) and correlates of attitudes indicates the need for care in the choice of particular frames. But such caution is always necessary in the use of question frames. As long as the frames reflect the actual language of political debate, framing effects must be considered substantively important variation rather than mere artifacts, and so cannot be invoked as a reason for rejecting frames as a means toward improved attitude measurement.

The fact that framed items can induce larger numbers of respondents to offer meaningful opinions -- in the sense that the opinions are at least stable -- on issues on which they might otherwise assert a casual don't know is a further practical value of using framed items.

Theoretical Implications

The first set of experiments produced strong evidence that different issue frames can induce different types of respondents to support a given issue -- though frame differences do not always have this effect. The second set of experiments showed that framed items enable most types of respondents to make more stable attitude reports but that such items also undermine the ability of extreme ideologues to do so. It thus appears that different individuals employ different methods of answering questions, and that use of framed questions affects these question-answering styles differently.

Both sets of findings support the central assertion of the question answering model, which is that surveys are not the occasion for revealing pre-formed opinions, but are instead the stimulus for pulling together one's ideas, as best one can, into a coherent survey response. The unevenness of the results obtained in these experiments indicate that innovations inspired by the question-answering approach will not likely bring rapid solutions to the many problems of measuring and understanding mass political attitudes. But even so, the approach has enjoyed some definite success in the experiments reported in this paper. It is, in addition, still the only theoretical approach that is capable even of addressing the types of phenomena examined here. One must therefore conclude that, despite the somewhat ragged results obtained in these experiments, the question-answering approach merits continued development and testing.

APPENDIX

Stealth bomber - - Communism frame

Here is a question about defense policy. There is much debate over whether the Air Force should build the new Stealth bomber. Some people are against this weapon. They say the Stealth is another costly high-tech weapon that will probably not work very well. Others believe we need the Stealth bomber. They say the only way to be safe from Russia and the other Communist bloc nations is to make sure we stay a lot stronger than they are. Do you believe the Air Force should go ahead with plans to build the Stealth bomber, or do you think that building the Stealth bomber would be a bad idea?

Stealth bomber - - immorality frame

Here is a question about defense policy. There is much debate over whether the Air Force should build the new Stealth bomber. Some people are against this weapon. They say the only way to stop the immorality of the nuclear arms race is for the U.S. to stop building up its nuclear forces. Others believe we need the Stealth bomber. They say our existing bomber forces are obsolete and need to be modernized. Do you believe the Air Force should go ahead with plans to build the Stealth bomber, or do you think that building the Stealth bomber would be a bad idea?

Oil drilling -- foreign dependency frame

There is a lot of talk these days about a plan to allow more drilling for oil on federal lands in Alaska. Some people are opposed to this drilling. They say the Alaskan wilderness should be preserved in its natural state for future generations. Others say the drilling is necessary because the U.S. needs new energy sources to avoid becoming dependent on foreign oil. What is your opinion? Do you favor or oppose drilling new oil fields on federal lands in Alaska?

Oil drilling -- lost jobs frame

There is a lot of talk these days about a plan to allow more drilling for oil on federal lands in Alaska. Some people are opposed to this drilling. They say there would be no need for this oil if we made better use of our existing energy resources. Others support the drilling. They say that without new sources of oil, working people will be hurt by higher gasoline prices and there will be further layoffs in American factories. What is your opinion? Do you favor or oppose drilling new oil fields on federal lands in Alaska?

Death Penalty -- Immoral frame

There is still much controversy about the death penalty in murder cases. Some people favor the death penalty because they believe it deters crime. Others oppose the death penalty because they believe killing another human being is always immoral, even the killing of someone who has committed murder. Do you favor or oppose the death penalty for the crime of murder?

Death Penalty -- unconstitutional frame

There is still much controversy about the death penalty in murder cases. Some people say that murder is so awful a crime that it deserves to be punished by death. Others oppose the death penalty. They say it is unconstitutional because it is "cruel and unusual punishment." Do you favor or oppose the death penalty for the crime of murder?

Contra aid -- communism frame

Here is another question about foreign policy. As you know, the U.S. has been giving aid to the Contras, a guerilla group that wants to overthrow the Communist government of Nicaragua. Some people say we should stop aid to the Contras because the U.S. has no business in the internal affairs of Nicaragua. Others think the U.S. should continue the aid because the Contras are freedom fighters trying to stop the spread of Communism in Central America. Would you like to see aid to the Contras in Nicaragua increased, decreased, or kept about the same?

Contra aid -- pro-democracy frame

Here is another question about foreign policy. As you know, the U.S. has been giving aid to the Contras, a guerilla group that wants to overthrow the Sandinista government of Nicaragua. Some people say we should stop aid to the Contras because the money could be better spent in the U.S. Others think the U.S. should continue the aid because the Contras are fighting to promote democracy in Central America. Would you like to see aid to the Contras in Nicaragua increased, decreased, or kept about the same?