

Frame of Reference and the Survey Response*

John Zaller
University of California, Los Angeles

and

Stanley Feldman
University of Kentucky

9-30-87

* This is a modified version of a paper prepared for delivery at the annual meetings of the American Political Science Association, September 3-6, 1987 in Chicago, Illinois. The data used in the paper were collected by the Board of Overseers of the National Election Studies under a grant from the National Science Foundation. Collecting these data made unusually heavy demands on the staff of the Institute for Social Research at the University of Michigan, where the study was conducted. In this connection we are grateful to ZoAnne Blackburn, Steve Pinney, and Santa Traugott, who supervised the study, and to the following members of the interview and coding departments: Heather Andrews, Valerie Austin, Linda Bailiff, Sylvia Baker, Lisa Bonet, Jim Buckmaster, Laurel Burke, Liesa Busse, Sujean Choi, Anmarie Currier, Jileen Devries, Jane Fisher, Doris Guthrie, Stan Hasper, Lori Johnson, Kurt Harju, Ann Haven, Lynn Hayes, Suzanne Hogan, Rachael Howe, Kelly Ladd, Yvonne Kaplan, Hedy Malki, Anna Moyer, Ann Munster, Judith Peebles, Louie Pettit, Ramon Randolph, Karen Smith, Kathryn Terrazas, Jeff Toliver, Marielle Weindorf. We also thank Barbara Geddes and Don Kinder for their very helpful comments on the initial draft of the paper. Finally, we would like to thank Hiroaki Minato for invaluable assistance in preparing the data for analysis. Neither these individuals nor institutions bear any responsibility for any errors of judgment or fact that may appear in this paper.

Abstract

Public opinion specialists customarily take the "attitude statements" given in response to survey questions as their fundamental units of analysis. This paper maintains that it is useful to disaggregate these attitude statements to a more primitive level of analysis, that of the "considerations" or "frames of reference" that underlie the attitude statements.

The paper is primarily concerned with understanding the massive response instability that occurs when questions are asked of the same people on two or more occasions. It argues that much of this instability may arise because people interpret the same question so differently in different interviews that they are, in effect, often answering different questions.

The data for investigating these issues come from the 1987 NES Pilot study. In a two-wave survey, this study asked respondents both to answer standard closed-ended items, and to discuss in open-ended fashion the issues raised in those items.

Most public opinion research assumes that the answers people give to survey questions are relatively direct measures of what the people actually believe. The view is that "if someone says he favors X, then he favors X." The weight of recent research, however, makes this conventional assumption increasingly dubious. If a respondent is asked the same question in a later interview,¹ or in a different context,² or in a slightly different form,³ he has an excellent chance of asserting "not X" rather than "X." This heavy dependence of survey responses on what ought to be accidental features of the interview process is more than simply a methodological curiosity; it is a continuing embarrassment to our notions of what public opinion is.

In light of this, some researchers are working toward a new conception of what survey responses represent.⁴ The emerging view is that citizens do not typically possess fixed attitudes toward political issues; rather, they seem to possess a mix of only partially consistent impressions and concerns. When asked a question in a survey, they must call to mind at least some of these opinion fragments and transform them into an appropriate survey response. The nature of this answer-generating process is seen as the key to understanding the response anomalies just described.

The present paper applies these new conceptions to the problem of response instability. It tries, that is, to explain why, in interviews only a few months apart, typically only about half of the general public will give the same answer both times.⁵ The paper begins with evidence confirming that the public, as Hochschild (1981) has argued from a restricted sample, often seems genuinely conflicted in its political attitudes. This is inferred from the fact that, when

¹ Converse (1964), Achen (1975), Dean and Moran (1977), Erikson (1979), Feldman (1985).

² Schuman and Presser (1981), Bishop, Oldendick and Tuchfarber (1985), Tourangeau and Rasinski (1986), Loftus, Feinberg, and Tanur (1985)

³ Schuman and Presser (1981), Schuman and Scott (1987).

⁴ A conference in Ann Arbor last summer brought together many of those working on this problem. See generally, Bishop, Oldendick and Tuchfarber (1982), Zaller (1984), Feldman (1985), Tourangeau and Rasinski (1986).

⁵ This claim applies to items having two substantive options, a middle response, and a don't know response.

invited to discuss the issues raised in standard policy items, people regularly bring up considerations on both sides of the question. The paper goes on to test specific models of how, in the face of internal conflict, people generate answers to survey questions. In particular, it tries to determine whether survey responses are determined by a single, dominant consideration, or whether people formulate answers by aggregating across numerous competing considerations. We argue that the available evidence, though not conclusive, is more consistent with the assumption of a single dominant consideration. More generally, we argue that individuals seem to understand the same question in different ways at different times, and that these differences in "frames of reference" powerfully affect the answers they give.

We cannot claim in this paper to have resolved the mystery of response instability. One point, however, does seem clear from our analysis: lurking beneath the tidy surface of the "attitude statements" to which people subscribe in opinion surveys, there exists a lively mental landscape, one characterized by much more internal conflict than most of the political behavior literature has heretofore recognized.⁶ Neither the nature nor importance of this conflict has yet been established. But careful investigation of it seems a most promising avenue for resolving the anomalies noted in the opening paragraph.

The principal data employed in this study are responses to open-ended probes that were asked in association with closed-ended questions about three domestic policy issues. A random half of a May, 1987, national survey was asked what ideas came to mind as they thought about key phrases in each of the three closed-ended questions; people responded to these probes just prior to answering each of the standard policy items. The other half of the national sample was asked, just after answering each closed-ended item, to describe exactly what was on their minds as they did so. The open-ended probes, along with the associated closed-ended questions, were repeated in a second interview approximately one month later. The data from these re-interviews enables us to gauge

⁶ See, however, Graber (1985).

the stability of both the policy attitudes and the considerations that presumably underlie them. All of these data were generated by the 1987 Pilot Study of the National Election Studies.

THEORETICAL BACKGROUND

In what, after more than two decades, is still the most important discussion of the nature of political attitudes, Converse (1964) argued that response instability arises in large part because individuals who lack meaningful attitudes on an issue nevertheless indulge interviewers by supplying answers -- answers chosen essentially at random. "[L]arge portions of an electorate," he suggested, "simply do not have meaningful beliefs, even on issues that have formed the basis for intense political controversy among elites for substantial periods of time (1964: 245)."

This conclusion has been challenged most forcefully by Achen (1975), who argued that individuals' "true attitudes" are overwhelmingly stable and that they appear to vary only because they are observed with substantial amounts of "measurement error." Such error arises, in his view, because of the inherent vagueness of all language, including that used in survey questions. Whereas Converse had, in effect, blamed respondents for having vague minds, Achen argues that response instability stems in significant part from vague questions -- vague in the sense that respondents have trouble figuring out exactly what they mean.

Both approaches to understanding response instability seem to us to have serious limitations. Converse's "non-attitudes" thesis, which assumed that anyone with a real attitude would be essentially perfectly stable, and that everyone else would respond to survey questions randomly, was an extreme claim intended to characterize mass attitudes in only certain limiting cases. Converse made this clear both in his initial statement of the model and in subsequent commentaries in which he refers to attitude crystallization as a continuous rather than a dichotomous concept (e.g., Converse and Markus, 1979). Still, the question remains: what exactly does it mean to have a crystallized attitude? No one has ever said.

The measurement error models of response instability seem to us equally underspecified at their theoretical core. When, as seems to be the case, measurement error constitutes two-thirds

or more of the variance of typical attitude items, one naturally wants a great deal of information about what this "error" consists of and how it has been generated. Yet the measurement model tradition has so far told us little more about the nature of measurement error than we know about the nature of attitude crystallization.

What is needed, we believe, is a new theoretical cut at the problem of response instability, one holding the promise of a more detailed account of how response instability occurs.

Jennifer Hochschild's study of citizens' attitudes toward equality may provide the basis for such a new approach. From depth interviews of 28 respondents, she found that people would, if asked to do so, readily answer fixed format questions, but that given the opportunity to talk,

... people do not make simple statements; they shade, modulate, deny, retract, or just grind to a halt in frustration. These manifestations of uncertainty are just as meaningful and interesting as the definitive statements of a belief system.

Hochschild particularly emphasizes the ambivalence of many of her interview subjects. This ambivalence frequently leads them to change their minds or contradict themselves-- which is to say, to give temporally unstable responses -- in the course of a single conversation. Consider her account of the attitudes of one of her subjects toward government income guarantees

Vincent Sartori cannot decide whether or not the government should guarantee incomes, because he cannot decide how much weight to give to the value of productivity. He believes that the rich are mostly undeserving and ... yet he is angry at "welfare cheats" who refuse to work... Caught between his desire for equality and his knowledge of existing injustice, on the one hand, and his fear that a guaranteed income will benefit even shirkers, on the other, he remains ambivalent about policies toward the poor.

The current debate over whether response instability arises from "vague minds" or "vague questions" seems quite irrelevant to Hochschild's observations. The reason for Vincent Sartori's vacillation is not that he has no real opinion, nor is it that he has a "true attitude" that Hochschild is unable to measure reliably. It is simply that he has conflicting opinions, or at least conflicting considerations, that lead him to give different responses at different times, depending on how he frames the issue in his own mind.

One can raise many objections to Hochschild's study -- that it is based on a non-representative sample, that she pays disproportionate attention to her more articulate

respondents, that her efforts to make people think critically about their attitudes may have created ambivalence where none had previously existed. Nonetheless, we are persuaded that her point about ambivalence represents an important theoretical insight. There is no reason to believe that individuals carry around in their heads fixed and perfectly coherent opinions about every question that survey researchers may happen to ask about. More likely, they carry around in their heads a mix of more or less consistent ideas and concerns that they can, if necessary, aggregate into responses to questions that are asked of them.

How exactly individuals go about formulating such answers is obviously an extremely difficult issue. Before beginning to grapple with it, we shall need to acquaint readers with the procedures that generated our data and with the basic contours of the data. Having done this, we shall specify and test some simple models of how exactly people formulate responses to survey questions.

THE DATA

The 1987 Pilot study conducted telephone interviews with 457 respondents in the May wave and 360 respondents in the June wave. These respondents had previously participated in the 1986 National Election Study. To insure greater representativeness than is typically achieved in panel studies, the Pilot study systematically oversampled respondents who scored at the lower end of the information scale in the 1986 survey; this reduced but did not quite eliminate underrepresentation of the less informed in the final Pilot sample (Zaller, 1987).

The policy items used in this study are the standard NES items concerning a guaranteed standard of living versus individual responsibility, aid to blacks versus individual responsibility, and cuts versus increases in the level of government services. They were converted from seven-point scales to simple dichotomies for use in the telephone survey.

The questionnaire had two forms. In form A, respondents were asked the several open-ended probes immediately after answering each of the three policy items. The exact form of these "retrospective" probes was:

Still thinking about the question you just answered, I'd like you to tell me what ideas came to mind as you were answering that question. Exactly what things went through your mind.

Are there any (other) reasons that you favor [the option just selected]?

Do you see any problems with [the option just selected]?

The first of these probes was designed to elicit a sort of "memory dump" which would reveal the considerations that were most important in determining respondents' answers. Prior work by Ericsson and Simon (1984) shows that such probes can work effectively if asked immediately after a given task has been carried out.

In form B, interviewers read the policy items in the standard way, but, without pausing for the respondent to answer, asked the respondent to "stop-and-think" about the question.⁷ In the case of the standard of living question, the exact probes were:

Before telling me how you feel about this, could you tell me what kinds of things come to mind when you think about government making sure that every person has a good standard of living? (Any others?)

Now, what comes to mind what you are thinking about letting each person get ahead on their own? (Any others?)

The interviewer then re-read the entire question and recorded the respondent's reply to it.⁸

Responses to the "retrospective" and "stop-and-think" probes should not be considered equivalent. The "retrospective" probes probably would not be expected to elicit an unbiased sampling of the full range of considerations in the individual's mind; they should, however, be a good indicator of why the individual answered the question as he did. The "stop-and-think" probes, on the other hand, ask for spontaneous reactions to various policy ideas. Since the individual has not yet committed himself to an answer, and may still be deciding which one to offer, responses to

⁷ This apt designation is the invention of Kathleen Knight,

⁸ In both the "retrospective" and "stop-and-think" condition, interviewers also transcribed any stray comments made by the respondent while listening to the question or deciding how to answer it. Requests to have the question repeated, as well as other kinds of non-substantive comments, were also recorded.

these probes may be expected to be more broadly representative of the range of ideas in the person's mind.

Interviewers attempted to write down as faithfully as possible everything that was said in response to the open-ended probes. Having listened to several interviews and afterwards examined what the interviewers wrote down, we felt that they did a good job of capturing respondents' exact language in cases when only one or two comments were made. However, some respondents spoke for a minute or more, making it impossible for interviewers to approach a verbatim transcription; here, the recorded comments are best regarded as thematic highlights of what the respondent actually said.

The transcribed comments were subjected to an elaborate classification scheme. As many as four comments were coded in connection with each probe, with each comment rated on several variables. The most important variable was "direction of comment," which indicated which side of the issue, if any, the remark tended to support. Although this variable noted ambivalence, confusion, and non-substantive concerns, three quarters of all comments had a clear directional thrust. Most of the remaining comments indicated a substantive concern, but failed to indicate clearly its evaluational thrust. The other key variable was the "frame of reference" code, a variable which included more than 140 categories and tried to capture the substantive content of each remark.⁹ The codes referred to general principles (e.g., equality, the role of government), individualism and the work ethic, the fairness of the economic system, particular population groups (e.g., blacks, the elderly), personal experience, and particular government programs.¹⁰

It is natural to wonder about the quality of the remarks people made. Did they say things that made sense? Or were they "just talking at random" because they had been asked to talk?

⁹ The coding scheme was derived empirically from pretests and modified as the need arose during the coding of the Pilot data.

¹⁰ About a tenth of all interviews were double-coded. Although exact reliability data are not available, it appears that there was a "difference" between coders on about 10 to 15 percent of all cases. The coding department regarded this difference rate as being within the normal range for open-ended material of this type.

Our coding scheme is our only means of addressing this difficult concern. The coding scheme is as elaborate as it is because people gave us answers that required this degree of elaboration. With relatively few exceptions, they did not simply recite the closed-ended questions back to the interviewers, but raised a variety of obviously pertinent concerns. Thus, we are convinced that most of what people said made sense. A much more difficult matter to assess is how much people cared about the various issues they raised. Discussion of that problem must, however, be deferred to our analysis of how exactly people transform their various considerations into answers to survey questions.

One other feature of our data must be noted. Although response instability is an unwanted feature of virtually all panel studies, there was relatively little response instability on the three items used in this study. Presumably, this resulted from the conjunction of an unusually short period between re-interviews and the requirement that people discuss their ideas in open-ended fashion. In any event, we have much less variance in our dependent variable than would ordinarily be present.

INITIAL RESULTS

Prior to undertaking this study, we were unsure how much people would be willing or able to say in response to our probes. On some readings of Converse's "non-attitudes" thesis, a sizeable segment of the public might have simply no reaction at all to many public issues. Our first and perhaps most solid finding, then, is that, despite our initial concern, most respondents did have a fair amount to say.

Figure 1 shows the distribution of substantive comments elicited by the stop-and-think probes. The average is about four codable comments per issue, with almost all respondents offering at least one substantive comment. Although, as shown in Table 1, the frequency of these substantive comments varies by individuals' level of political information, even the least informed respondents generally offered one or two substantive remarks. The proportion of respondents saying "no opinion" in the stop-and-think condition was negligible, so these rates are

based on nearly the whole sample. (The higher total number of remarks on the race item is probably artifactual; respondents were probed about three phrases in that question and about two in the other questions.)

Insert Figure 1 and Table 1 about here

Table 2 assesses the level of conflict in what people had to say. Conflict here is measured by the number of opposing remarks that could be paired against one another. If a respondent made, for example, two comments with a liberal thrust and two with a conservative thrust,¹¹ his score on the conflict scale would be two. If he made three (or more) on one side of the issue and two on the other, the conflict score would still be two since the number of comments that could be matched would remain two.

Insert Table 2 about here

As can be seen, an average of about 30 percent of the sample expressed minimally conflicting evaluations of an issue within a single wave of the survey, and an additional 10 percent or so counterbalanced two comments favoring one side of the issue with two or more comments favoring the other side. Since these remarks were manifestly evaluative and were made within a few seconds of one another -- often in response to the same probe -- most respondents were probably consciously aware of the inconsistency.

When the same conflict measure is calculated for comments made across both waves of the survey, we find, as would be expected, much higher levels of internal conflict. As Table 2 shows, an average of about two-thirds of all respondents offer conflicting evaluations. That is, a majority of them volunteer at least one reason for coming down on each side of the given policy issue.

¹¹ Here as elsewhere, we describe an individual's comments as having a liberal or conservative thrust without meaning to imply that the comment was necessarily ideological in flavor. We mean only that the comment seemed to incline the respondent toward the liberal (or the conservative) response option.

Figure 1

Frequency of Substantive Comments on Open-ended Probes

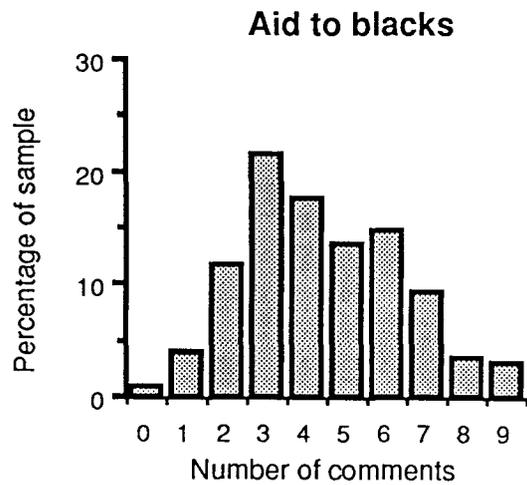
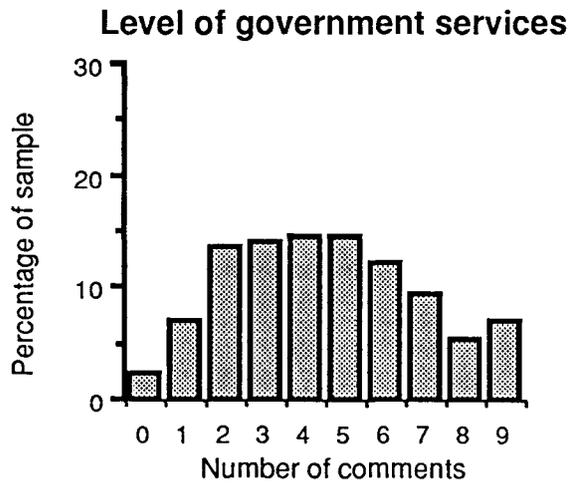


Table 1

**The Effect of Political Information
on Volume of Open-ended comments**

	<u>LEVEL OF POLITICAL INFORMATION</u>					
	<u>AVE.</u>	<u>Low</u>	<u>—</u>	<u>Medium</u>	<u>—</u>	<u>High</u>
Standard of Living	3.4 *	2.6	3.3	3.7	4.0	3.7
Government Services	4.2	2.6	3.9	4.4	5.2	5.4
Aid to Blacks	4.5	3.4	4.2	4.6	5.5	4.6

* Cell entries give the average number of discrete substantive remarks in the given cell.

Table 2

**Level of Conflict in Remarks
About Political Issues**

CONFLICT WITHIN A SINGLE WAVE OF THE SURVEY

	<u>GUARANTEED LIVING STANDARD</u>	<u>GOVERNMENT SERVICES</u>	<u>AID TO BLACKS</u>
No. of conflicting frames used*			
0	63.2%†	57.3	45.7
1	27.8	29.1	37.9
2	8.0	10.5	12.9
3 +	.9	3.2	3.4
	100%	100%	100%

CONFLICT ACROSS TWO WAVES OF THE SURVEY

	<u>GUARANTEED LIVING STANDARD</u>	<u>GOVERNMENT SERVICES</u>	<u>AID TO BLACKS</u>
No. of conflicting frames used*			
0	36.6%†	38.7	20.9
1	27.9	20.9	30.6
2	22.1	21.8	25.4
3 +	13.4	18.7	23.1
	100%	100%	100%

† Cell entries are percent of sample in "stop and think" condition at each given level of conflict.

These findings are obviously consistent with Hochschild's portrait of a citizenry that sees more than one side to issues but does not always resolve -- and may not even recognize -- its own contradictory impulses. Still, these findings constitute only a first cut at the problem. How stable are individuals' evaluational remarks? How are the evaluational remarks related to choices on the closed-ended attitude items? How should they affect our understanding of what closed-ended survey responses represent? We address these more substantively interesting questions below.

THE STABILITY OF QUESTION FRAMES

When asked in the May interview about the proper level of government services, a school teacher stated emphatically that he favored more federal spending. The country was facing an educational crisis, he said, and more education expenditures were drastically needed. Any cuts in federal spending would inevitably reduce the already inadequate amounts of money available for education. Just a month later, however, the same individual said that he favored cuts in government spending. Government was too big, he contended, and something had to be done to cut it back. There was no reference at all to the "educational crisis" that had preoccupied him in the earlier interview.¹²

Opinion researchers have long known that different people can answer identical questions as if the questions concerned completely different subjects. What our vignette of the vacillating teacher suggests is that the same people can answer the same question at different times as if it were two different questions. This presumably happens because the question strikes them differently -- and hence means something different -- in the two interview settings. A question about government services may raise the spectre of special interests and bloated bureaucracies at one interview, and an image of education, social security, and air safety at another. A question about a guaranteed standard of living may suggest "welfare cheats" at one point and the plight of displaced

¹² We would like to present verbatim transcriptions of what this and other respondents said. However, the Board of Overseers of the NES strictly prohibits such use of the protocols. In consultation with the human subjects protection committee at the University of Michigan, they have determined that any publication of the raw protocols would be an invasion of the respondents' right to privacy.

steel workers at another. Thus, the same person can give opposing "survey responses" without changing his underlying attitudes, if he has shifted the frame of reference by which he understands the question.

Our data were collected with a specific view to detecting just such changes in frames of reference. In examining the data, however, the reader should bear in mind their inherent limitations. Citizens may express the same basic idea in different language, different interviewers may happen to catch different phrases across the two waves, the coders may "read" the comments in one or more plausible ways, and the coding scheme itself may be flawed.¹³ The effect of these and other likely sources of error would be a bias toward finding more changes in frames of reference than actually occur.

Notwithstanding all of this, the data presented in Table 3 create the strong impression that questions really do strike people differently at different times -- even people whose responses to the closed-ended questions remain stable.¹⁴ Consider respondent A. His initial reaction to the notion of a guaranteed standard of living was that it was inconsistent with American ideals; yet he also expressed reservations about letting individuals get ahead on their own, saying that some people need special help and that society has an obligation to help the needy. He then went on to express concern about the unfairness of supporting those who don't work, concluding with the ambivalent statement that although individuals should get ahead on their own, some are unable to do so.¹⁵ One gets the impression that, although this person came down on the side of individual responsibility in the first interview, he might well have answered the question the other way. This ambivalence is, however, completely absent in the second interview. The respondent gives six reasons why individuals ought to get ahead on their own, including a conservative reference to

¹³ It is, we should add, far from obvious that these kinds of potential error are more serious than the types of error known to be associated with standard fixed-format questionnaires and with depth interviews.

¹⁴ The cases in Table 3 were selected randomly, subject to the condition that we wanted two who were stable and two who were not.

¹⁵ For a fuller description of the codes on which this commentary is based, see the ICPSR codebook for the study.

the value of education in insuring equal opportunity. Having given no hint of his earlier feelings of obligation toward the needy, his conservative choice on the closed-ended item is now entirely expected.

Insert Table 3 about here

Respondent A, then, is far less stable in his reaction to the guaranteed standard of living question than his stable closed-ended responses would suggest. He has gone from being an ambivalent conservative on this issue to being a confident conservative. Thus, one would not be surprised if, over repeated trials, he occasionally expressed support for a guaranteed living standard. Certainly he has at least some impulse in that direction.

Consider next respondent C. He seems relatively stable in his basic reaction to the issue even though his attitude on the fixed item changes from conservative to liberal. He is ambivalent at both interviews to approximately the same degree and on the basis of similar considerations. One can imagine that he may have, in effect, tossed a mental coin in deciding how to answer the fixed question -- not because, as Converse's non-attitudes thesis suggests, he had no meaningful substantive reaction to the issue (clearly, he did), and not because, as the measurement error models suggest, he wasn't quite sure what the question was asking (he saw it the same way at both interviews), but because he was undecided between largely stable but conflicting impulses.

THE STABILITY OF CLOSED-ENDED "SURVEY RESPONSES"

We have examined the stability of question frames partly because it is interesting to know whether people interpret questions the same way each time they answer them, and partly because we think it is likely to be related to the stability of responses to the closed-ended questions. We turn now to a test of the latter expectation.

In conducting this test, we wanted to measure the ideas that were important to the individual at the moment of answering the closed-ended questions. Responses to the "retrospective" probes seem ideally suited to this purpose. Although, as noted earlier, they probably tap a narrower

Table 3

Content of Open-ended responses on Job guarantees question

Could you tell me what kinds of things come to mind when you think about...

<u>(FIRST WAVE)</u>		<u>(SECOND WAVE)</u>	
<i>...government making sure that every person has a good standard of living?</i>		<i>... letting each person get ahead on their own?</i>	
<i>... government making sure that each person has a good standard of living?</i>		<i>... letting each person get ahead on their own?</i>	
<u>Respondent #A</u>	(STABLE CONSERVATIVE)		
110. Idea is un-American* (Cons.)**	180. Some people need help (Lib.)	158. Tax burden too great (Con.)	140. Individualism/work ethic (Con.)
	161. Duty to help the needy (Lib.)	110. Idea is un-American (Con.)	133. Equal opportunity exists for all (Con.)
	136. Unfair if some don't work (Con.)	346. Program/food stamps (Con.)	344. Program/educ. (Con.)
	145. All should make it alone but some need help (ETN)*		
<u>Respondent #B</u>	(STABLE CONSERVATIVE)		
158. Tax burden too great (Con.)	344. Program/Educ. (Lib.)	344. Program/educ. (Lib.)	140. Individualism/work ethic (Con.)
156. Govt. red tape (Con.)	344. Program/Educ. (Lib.)	219. Group ref. to middle class (Con)	142. Shiftless people deserve fate (Con.)
144. Value of competition (Con.)	110. Idea is un-American (Con.)	152. Limited gov't. (Con.)	147. Motivation to work (Con.)
		140. Individualism/work ethic (Con.)	

* A full description of these codes can be found in the ICPSR codebook for the 1987 Pilot Study.

** Refers to direction thrust of comment; ETN indicates that the evaluational thrust of the remark, if any, was not obvious.

TABLE CONTINUES ON NEXT PAGE

(TABLE 3 CONTINUED)

<u>(FIRST WAVE)</u>		<u>(SECOND WAVE)</u>	
<i>...government making sure that every person has a good standard of living?</i>	<i>... letting each person get ahead on their own?</i>	<i>... government making sure that each person has a good standard of living?</i>	<i>... letting each person get ahead on their own?</i>
<u>Respondent #C</u>	(UNSTABLE, Conservative to Liberal)		
140. Individualism/ work ethic (Cons.)	140. Individualism work ethic (Con.)	151. Gov't. must insure equal opp. (Lib.)	140. Individualism work ethic (Con.)
161. Duty to help the needy (Lib.)			
<u>Respondent #D</u>	(UNSTABLE, Liberal to Conservative)		
344. Program/educ. (Lib.)	140. Individualism/ work ethic (ETN)	150. Idea of welfare state (Lib.)	111. Fairness of Amer. system (Con.)
357. Program/ housing (Lib.)		348. Program/ health (Lib.)	
356. Program/welfare (Lib.)		143. Work & welfare (Lib.)	

range of ideas than the "stop and think" probes, they are more likely to elicit the ideas at the "top of the head" at the moment of answering the closed-ended question.

The actual measure of "frame stability" is a count of how often individuals mentioned the same or closely similar substantive concerns (as captured by the coding scheme described earlier) across the two waves of the interviews.¹⁶ Although some codes imply a particular evaluational thrust, the measure itself pays no explicit attention to directional thrust, but only to content. Thus if, for example, someone mentioned health and education programs as reasons for favoring more government spending in the first wave, and then cited health and education programs as arguments against government spending in the second wave, he would be counted as having a stable frame of reference.¹⁷

The effect of having a stable frame of reference is shown in Table 4. For all three issues, repetition of several substantive comments across waves is associated with markedly greater stability on the closed-ended items -- actually, perfect stability.¹⁸ Unfortunately, the limited amount of response instability in these data prevents us from getting a good fix on the power of frame stability as a determinant of instability.¹⁹

¹⁶ We used only responses to the first two retrospective probes; the third probe was intended to turn attention away from what was on the individual's mind at the moment of answering by probing for "difficulties" with the recommended policy.

¹⁷ In fact few such changes occur. The vast majority of respondents who change the overall evaluational thrust of their comments refer to quite different substantive concerns in doing so.

¹⁸ The measure of frame stability was constructed as follows: An individual received two points if his first comment on both interviews fell into the same substantive category. Otherwise, he received one point each time a code used in discussing an issue in the first interview was used in discussing that same issue in the second interview. Higher scores may be interpreted as indicating a greater likelihood that the same or nearly the same frame was used on both interviews.

To increase the chance that the similarities of comments would be detected, the "frame of reference" variables were recoded as follows: (132=146) (140, 141, 142=148) (145=153) (161, 162=180) (221=222) (230, 231=232) (235=236) (238=239) (241=246) (243=244) (322=336) (323=356) 324=343) (325=326) (340=360) (349=355) (440=441) (510=511) (171=237) (157=312) (210=354) (213=341) (215=356) (250=340). The affect of these recodes is to slightly improve the distribution of cases in Table 4; it does not greatly effect the relationship between frame stability and response stability.

¹⁹ On the stop-and-think side, frame stability had little effect on response stability. This is the first of many indications to come that the retrospective and stop-and-think probes measured different things.

Insert Table 4 about here

There is one other reason for considering this test a conservative one. Owing to the many potential sources of error discussed earlier, there were surely some respondents who used essentially the same frames of reference in both surveys but failed to be counted by our rather stringent test of frame stability. Thus, the test probably underestimates the extent to which people employed the same issue frame in successive interviews.²⁰

Table 5 gives further evidence on the importance of stable frames on the stability of responses to the closed-ended items. It shows that the stability of closed-ended responses varies as a function of the consistency and stability of individuals' evaluational remarks. The consistency-stability measure was generated as follows: First we calculated the net proportion of evaluational remarks that were "liberal" or "conservative," a measure that must then vary between -1 to +1. Next we summed these two measures across waves and took the absolute value. The resulting measure gives high scores to individuals whose evaluational remarks were consistently liberal or consistently liberal on both waves of the survey, and low scores to people whose evaluational remarks were either conflicted or vacillating. Table 6 shows that this measure of "stable consistency" predicts stability on the closed-ended items quite well -- and in two of the cases, perfectly well. The only poor performance concerns the services item on the "stop and think" side, a problem that appears to be due to coding error.²¹

²⁰ The evaluational remarks of people in the zero cell of Table 4 are by no means devoid of stable evaluational content. To show this, we created a summary measure of the directional thrust for each respondent; this measure predicts responses to the fixed items on the second wave almost as well for the subsample of people scoring zero on the "frame stability" measure as it does for the entire sample.

²¹ Since we have measures of both what people said on closed-ended items and how they justified their answers, we can check for consistency. That is, we can check whether people who make overwhelmingly liberal or overwhelmingly conservative remarks on the probes are always coded as having given the appropriate closed-ended response. The answer is that in about 5 percent of all cases, they are not. These seem clearly to be errors. Either the individual misspoke in giving the closed-ended response, or the interviewer recorded it incorrectly. (The open-ended remarks are too unambiguous to be in error.) In any event, such people usually "return" to their appropriate response in the next survey, which creates the effect shown for the services item on the lower portion of Table 6. When these cases of error are screened out, several more of the high

Table 4
Stable Frames of Reference
and Response Stability

No. of identical frames used*	<u>GUARANTEED LIVING STANDARD</u>	<u>GOVERNMENT SERVICES</u>	<u>AID TO MINORITIES</u>
	N	N	N
0	.79† (47)	.73 (45)	.89 (54)
1	.95 (22)	.94 (17)	.95 (21)
2	.92 (13)	.94 (17)	.83 (23)
3 +	1.00 (13)	1.00 (11)	1.00 (12)

* The measure is a count of the number of times a specific content code was repeated across the "retrospective" probes of the May and June waves of the survey.

† Cell entries are proportion of respondents who were stable in their answers to the fixed choice items between the two waves of the survey.

Insert Table 5 about here

Although the finding that stable frames are associated with stable attitude statements is clearly an important one, stable frames are more the exception than the rule. Indeed, even when coded strictly in terms of directional thrust (which is less subject to the various sources of error we have described) people's evaluational comments are typically both inconsistent and unstable. The evidence on instability is contained in Table 6. The net proportion of remarks having a liberal or conservative thrust correlated across waves of the survey at about .38 on the stop-and-think side and about .45 on the retrospective side.²² As Table 6 also shows, responses to the closed-ended items are, in every case, somewhat more stable than the considerations that supposedly underlie them. This is an interesting and revealing result to which we shall return below.

Insert Table 6 about here

ALTERNATIVE MODELS OF THE SURVEY RESPONSE

Our basic theoretical posture, as we have indicated, is that individuals can think about issues via any of several frames of reference, some of which may have contradictory implications for how they respond to the issue. (This part of our argument is scarcely new to us. The terms "schema," "interpretive perspective," "consideration," and "stereotype" are all examples of concepts that have been used by researchers in other fields to convey the notion that people have at their disposal multiple, organized points of view.²³) When people are asked a survey question, they must somehow convert these competing frames -- or *considerations*, as we shall often call

consistency cells move to a level of perfect stability, including the problematic services item. We intend to follow up on this unique opportunity to check for genuinely random error in later versions of this paper.

²² The measure used in this test is discussed in detail below.

²³ Although sharing much common meaning, each of these terms also has its own technical meaning. We make no attempt to distinguish among the competing concepts because we cannot see that our data would give us any empirical basis for preferring one over the others.

Table 5

The Effect of Stable and Consistent
Frames On Response Stability

RETROSPECTIVE PROBES

Consistency of frames used*	<u>GUARANTEED</u>	<u>GOVERNMENT</u>	<u>AID TO</u>
	<u>LIVING STANDARD</u>	<u>SERVICES</u>	<u>BLACKS</u>
	<u>N</u>	<u>N</u>	<u>N</u>
0	.50† (8)	.66 (12)	.85 (13)
1	.76 (17)	.62 (13)	.79 (19)
2	.95 (20)	.80 (20)	.88 (17)
3	.92 (25)	.96 (28)	.83 (18)
4	.96 (25)	1.00 (17)	1.00 (43)

"STOP AND THINK" PROBES

Consistency of frames used*	<u>GUARANTEED</u>	<u>GOVERNMENT</u>	<u>AID TO</u>
	<u>LIVING STANDARD</u>	<u>SERVICES</u>	<u>BLACKS</u>
	<u>N</u>	<u>N</u>	<u>N</u>
0	.68†(19)	.71 (24)	.70 (23)
1	.73 (45)	.71 (40)	.82 (39)
2	.75 (32)	.83 (41)	.85 (39)
3	.75 (28)	.85 (26)	.82 (35)
4	.93 (27)	.75 (8)	.90 (29)

* Consistency is measured by summing the absolute value of the consistency measures across each wave of the survey. The formula for creating the measure is as follows:
COMPUTE CONSIST=RND(ABS(Consist1+Consist2)*2))

† Cell entries are proportion of respondents who were stable in their responses to the given closed-ended item from wave one to wave two.

Table 6
Correlations Among Summary Evaluations
And Closed-ended Responses

	<u>Evaluations1</u>	<u>Evaluations2</u>	<u>Services1</u>	<u>Services2</u>
<u>Net Evaluations1</u>	1.00	.29*	.28	.28
<u>Net Evaluations2</u>	.35	1.00	.29	.32
<u>Gov't Services1</u>	.83	.47	1.00	.48
<u>Gov't Services2</u>	.49	.63	.58	1.00

	<u>Evaluations1</u>	<u>Evaluations2</u>	<u>Gov't jobs1</u>	<u>Gov't jobs2</u>
<u>Net Evaluations1</u>	1.00	.32	.39	.35
<u>Net Evaluations2</u>	.45	1.00	.33	.42
<u>Gov't jobs1</u>	.66	.48	1.00	.46
<u>Gov't jobs2</u>	.46	.58	.69	1.00

	<u>Evaluations1</u>	<u>Evaluations2</u>	<u>Black Aid1</u>	<u>Black Aid2</u>
<u>Net Evaluations1</u>	1.00	.53	.55	.51
<u>Net Evaluations2</u>	.56	1.00	.48	.48
<u>Aid to blacks1</u>	.66	.57	1.00	.63
<u>Aid to blacks2</u>	.64	.74	.77	1.00

* Cell entries are correlations between the indicated variables. Entries below the main diagonal refer to data from the retrospective probes; entries above the diagonal refer to data from the stop-and-think probes.

them -- into a choice among options offered by the closed-ended question. We now begin to consider the difficult question of how this conversion process might occur. Do individuals, for example, sample a wide range of considerations, sum across them, and answer on this basis? Or do they simply respond off the "top of their heads" -- that is, on the basis of the first consideration that occurred to them?

Either of these rules for transforming considerations into opinion statements is, or any of several possible hybrids, is quite plausible. Stanley Kelley (1983), for example, argues that voters decide which candidate to support by summing up all of their "likes" and "dislikes" about each party and presidential candidate and choosing the one with the highest net total. He considers but rejects the possibility that voters make this decision on the basis of a weighted sum.

Studies of social cognition, on the other hand, have repeatedly shown that the introduction or emphasis of a single piece of information -- such as the fact that a particular person is a woman or a lawyer -- can greatly affect subsequent expressions of opinion. Reviewing a variety of such evidence, Taylor and Fiske (1978) suggest that many people make social judgments by seizing on

"a single, sufficient and salient explanation... often the first satisfactory one that comes along... [I]nstead of employing base rate or consensus information logically, people are more often influenced by a single, colorful piece of case history evidence.... Instead of reviewing all the evidence that bears upon a particular problem, people frequently use the information which is most salient or available to them, that is, that which is most easily brought to mind. (p. 251)

Work on framing effects by Tversky and Kahneman (1982) likewise supports the view that individuals are often inordinately influenced by a single, dominant frame.

There is, of course, no reason to believe that people use the same decision procedures in all situations. Their decision processes may approximate the calculation of weighted sums when important matters are at stake, but individuals may still respond to questions off "the top of the head" by using a single, dominant consideration when little time is available for a decision, or when the matter is as inconsequential as answering survey questions (Taylor and Fiske, 1978).

Our data turn out to provide us more leverage than one might suspect in choosing between these various possibilities.

We begin with a simple but, we hope, revealing test. We will use two variables from the first wave of the survey -- responses to the closed-ended item, and the proportion of open-ended remarks favoring one or the other side of the item²⁴ -- to attempt to predict responses to the closed-ended item on the second wave of the survey. Consider several ideal-typical possibilities. Suppose, first, that individuals respond to survey questions by calculating a net score across all of the considerations that occur to them (the Kelley model, above). If this model were correct, the summary evaluations measure would contain much more information about how the individual would respond to the wave two item than would the wave one item. The wave one item is, after all, a dichotomy that could convey no hint of whether the choice was an easy or a difficult one.²⁵ The evaluational measure, by contrast, would reveal not only the net direction of the person's feelings, but how one-sided they were. It would, in other words, distinguish between the person who had five considerations favoring the given response and none opposing it, and the person who was ambivalent between sets of conflicting considerations. Thus in the regression test we envision, the evaluational measure would dominate the wave one item in predicting the wave two response.

Now consider what would happen if people were answering on the basis of the first idea that comes to mind, or alternatively, the single most important idea that comes to mind. The key idea in this "dominant consideration" model is that, although people might have in their minds numerous, conflicting considerations, and might be quite willing to talk about them, they would really care about only one of them. Hence they would engage in no "averaging." If this were the case, the evaluations measure would contain no more information about responses to the wave two item than would the wave one item itself. To be sure, information about the dominant consideration would be contained in the wave one evaluational measure, but so would information about numerous lesser considerations which have no effect on answering the question. The wave

²⁴ We experimented with several ways of combining the evaluational remarks into a summary measure. The one described in the text outperformed all others in predicting relevant criterion variables.

²⁵ The closed-ended items were converted to a dichotomous form for telephone administration. An "it depends" response was accepted if volunteered.

one closed-ended response, by contrast, would contain the information about the dominant consideration and would not contain information about irrelevant considerations. To the extent, therefore, that there were information about the individual's likely wave two response in the wave one material, it ought to be much more clearly present in the wave one response. And if that were true, the wave one item would dominate the evaluational measure as a predictor of the wave two item.²⁶

Table 7 presents evidence on the relative impact of the evaluations measure and the wave one item in predicting responses to the wave two items. All variables in the table are scored from -1 to +1, so the unstandardized regression coefficients are readily interpretable. For the half of the sample responding to the retrospective probes, the evidence heavily favors the dominant considerations model. For all six tests in column one, the wave one item has a large and statistically significant coefficient, while for five of the six tests, the evaluations measure fails to achieve statistical significance. In the final case, the impact of the evaluations measure is still less than that of the wave one item.

Insert Table 7 about here

For the half of the sample responding to "stop and think" probes, the evidence is mixed but still leans toward the dominant considerations model. Both the evaluational measures and the wave one items achieve statistically significant coefficients. But in all cases, the coefficients for the wave one items are larger than those for the evaluations scale.

Taken altogether, these results are more favorable to the dominant considerations model. In no case does the coefficient for the wave one item approach zero, as it ought to do if people were

²⁶ This discussion implicitly assumes that each "evaluational remark" that is counted as part of the summary evaluations measure represents exactly one "consideration." This is obviously a most unlikely assumption. But by converting the initial summary index to a proportional index, we partially control for individual differences in talkativeness. Also, it is reasonable to assume that if individuals elaborate on the considerations that occur to them -- so that the same basic consideration is counted two or more times -- they are more likely to elaborate on considerations that are more important. Hence the final measure ends up giving greater weight to considerations that are more important to the individual.

Figure 7
**Coefficients for Test of
Alternative Response Mechanisms**

	RETROSPECTIVE STOP-AND-THINK	
	<u>b/(se)</u>	<u>b/(se)</u>
Guaranteed Living Standard		
<u>Evaluations1</u>	.001* (.11)	.285 (.10)
<u>Gov't Item1</u>	.689 (.10)	.393 (.07)
<u>Evaluations2</u>	.170** (.11)	.223 (.10)
<u>Gov't Item2</u>	.607 (.09)	.382 (.07)
 Level of Government Services		
<u>Evaluations1</u>	.047 (.16)	.201 (.10)
<u>Services Item1</u>	.529 (.14)	.423 (.07)
<u>Evaluations2</u>	.267 (.14)	.207 (.10)
<u>Services Item2</u>	.487 (.10)	.433 (.07)
 Special help to blacks		
<u>Evaluations1</u>	.283 (.10)	.341 (.10)
<u>Blacks Item1</u>	.624 (.08)	.485 (.07)
<u>Evaluations2</u>	.011 (.11)	.348 (.10)
<u>Blacks Item2</u>	.769 (.09)	.537 (.07)

* Dependent variable is the wave two closed ended item

** Dependent variable is the wave one closed ended item

engaged in some kind of averaging process. But in five of the six cases on the retrospective side, the evaluations coefficient approaches zero, and in all cases on the stop-and think side these coefficients, though not zero, are smaller than the coefficients for the wave one items.²⁷

The fact that the evidence favors the dominant considerations model more strongly on the "retrospective" than on the "stop and think" side is quite consistent with the likely effects of the two kinds of probes. On the "stop and think" side, individuals were encouraged to think more broadly about the issue than they might ordinarily have done, and to do so before answering the question. This would tend to inhibit the expression of "top of the head" answers dominated by the first idea that came to mind, and to encourage the integration of multiple frames of reference. On the retrospective side, no such inducements to systematic decision-making were present. Thus according to both theory and data, more averaging occurred on the "stop and think" side.²⁸

These results, if interpreted correctly, indicate that people may normally answer survey questions "off the top of their heads," but that given appropriate inducements, they can take into account a broader range of considerations in formulating survey responses than they customarily do.

Happily, this interpretation has implications that are, at least in principle, testable. If responses to the closed-ended items on the "stop and think" side have really been based on a broader range of considerations, they ought to be more reliable than those on the "retrospective" side. In other words, the item responses on the "stop and think" side ought to be more stable over time, and more strongly correlated with relevant criterion variables (such as ideological self-designation).

²⁷ From a perusal of the zero-order relationships in Table 5, one certainly would not have expected the evaluations measure from the stop-and-think side to outperform its retrospective counterpart. The fact that it has done so underscores the point that the evaluations measure on the retrospective side mainly measures the effect of a single, dominant consideration.

²⁸ Unfortunately, this difference across forms might have been produced by the presumably greater tendency of people on the "retrospective" side to justify the answer just given by listing a long string of considerations consistent with it.

Unfortunately, our data do not permit a strong test of this implication (though we are currently engaged in a study that will permit such a test). The principal problem is that, owing to a failure of foresight on our part, the questionnaire induced much lower rates of "don't know" on the "stop and think" side. All else equal, this would have tended to produce more reliable responses on the "retrospective" side, because respondents there would represent a more selected and informed group.

Some reliability comparisons are nonetheless worth examining. To the extent that they turn out to be higher on the "stop and think" side, it would be despite, rather than because of, the differential don't know rates across halves of the study.²⁹

Table 8 presents correlations between the three closed-ended items and two measures of ideology, scores on the NES equalitarianism scale and self-placement on the seven-point liberalism-conservatism scale. As can be seen, these correlations are consistently higher for highly informed respondents on the "stop and think" side, and consistently lower for low information respondents on the "stop and think" side.³⁰ The former difference runs against the grain of the artifact and is therefore likely to represent an actual effect of the "stop and think" probes; the later difference, which is the reverse of our expectations, could have been produced by the artifact problem we described.³¹ (We are currently undertaking a new study to test our expectations about the effect of the "stop and think" probes on the reliability of closed-ended items.)

Insert Table 8 about here

²⁹ Contrary to expectation but consistent with the artifactual problem, response stability was slightly higher on the "retrospective" side.

³⁰ Paired difference of means tests in the low, middle and high information groups are statistically significant at levels of .1, .01, and .0005.

³¹ The artifact has sizeable effects in the low information groups, producing cells that are 30 to 50 percent higher on the "stop and think" side, and a fairly small effect on the high information group. It is possible that the "stop and think" actually does depress reliability among the poorly informed, perhaps because, for them, the greater thought induced by the probes leads to confusion.

TABLE 8
The Effect of "Stop & Think" on
Ideological Constraint

	<u>STANDARD OF LIVING</u>		<u>GOVERNMENT SERVICES</u>		<u>AID TO BLACKS</u>	
	<u>Top of Head</u>	<u>Stop & Think</u>	<u>Top of Head</u>	<u>Stop & Think</u>	<u>Top of Head</u>	<u>Stop & Think</u>
<u>LOW INFORMATION</u>						
Ideological Self-Identification	.17*	-.17	.30 [†]	-.17	-.02	.01
Individualism scale	.39 [†]	-.11	.05	.17	.28 [†]	.28 [†]
<u>MIDDLE INFORMATION</u>						
Ideological Self-Identification	.30	.16	.17	-.17	.13	.17
Individualism scale	.25 [†]	-.14	.40 [†]	.11	.29 [†]	-.09
<u>HIGH INFORMATION</u>						
Ideological Self-Identification	.42 [†]	.51 [†]	.26 [†]	.54 [†]	.43 [†]	.68 [†]
Individualism scale	.30 [†]	.46 [†]	.29 [†]	.48 [†]	.38 [†]	.57 [†]

* Cell entries are correlations between the scale on the left and the indicated item.

[†] Indicates that correlation is significant at level of .05 or less.

Three other pieces of evidence tend to favor the dominant considerations model. The first is that, as we saw earlier, individuals who raise the same substantive concern on both waves of the survey are extremely stable in their responses to the closed-ended policy items. If individuals were simply averaging across all of the considerations that came to mind, the stability of any particular consideration would not be especially consequential.

The second piece of evidence is also derived from the retrospective probes. As noted earlier, there were three separate retrospective probes, the last of which specifically asks whether the individual sees any problems with the policy option he has just selected. When the comments elicited from this probe are tallied separately from the others, they turn out to have virtually no empirical power to predict responses to the wave two closed-ended items. This suggests that individuals have some evaluational material near the tops of their minds to which they give no weight in formulating responses to closed-ended items, which is consistent with the dominant considerations model.³²

Finally, recall from Table 6 that responses to the closed-ended items were more stable over time than the evaluational measures. If the closed-ended items depend only on the stability of a dominant consideration, but the evaluational measures depend on the stability of the dominant consideration plus the stability of other considerations, then the item would be expected to be somewhat more stable than the scale.

We believe that the overall pattern of this evidence leans in the direction of a "dominant consideration" or "top of the head" model of the survey response and against an averaging model. Although it is that none of the evidence is conclusive, it seems all to run in the same direction. Obviously, however, much more and clearer evidence is needed before reaching any strong conclusions on this matter.

³² It is possible, however, that material elicited by a probe which specifically asks for countervailing considerations may be less important than material that is volunteered spontaneously.

If, however, it were the case that a single, dominant consideration determines most people's attitude statements, what mechanisms might enable it to do so? Our speculation is that soon after individuals read a question, they decide what they are being asked about. If, for example, the question refers to guaranteed living standards, they may, in effect, say to themselves, "Oh, this is a question about welfare cheats," or "Oh, this is a question about government keeping interest rates down so people can find jobs." Once such an interpretation is made, people's answers may be largely determined. They may go on to recognize doubts about their chosen alternative -- noting, for example, that some people may really need government help, or that interest rates are hard to control -- but as long as they continue thinking about the issue as a question about "welfare cheats" or about "exorbitant interest rates," their final attitude statement will remain firm.³³

DISCUSSION

The model of the survey response that has emerged from this study is disarmingly simple. It claims that the typical individual is capable of viewing political issues from several points of view, some of them in conflict with one another. The most important of these points of view -- or frames, or considerations, or schema (depending on one's theoretical tastes)-- that comes into an individual's consciousness at the moment of answering a survey question is that one that is likely to determine the answer to that question. To the extent that these frames are stable, survey responses are likely to be stable as well; when frames change, responses to the closed-ended responses are likely to change, too. There is, however, no ironclad necessity in this. Some individuals (e.g., respondent C from Table 3) may have stable but deeply ambivalent responses to issues, and this can lead to unstable responses to closed-ended questions. In other cases, closed-ended responses may remain stable even though the considerations underlying them change a great deal (e.g., respondent A in Table 3).

³³ Tourangeau and Rasinski (1986) propose a four-step question-answering process: people interpret the question, canvass their memories for relevant considerations, formulate an answer, and map that answer onto the original question. Our suggestion is that there may be only one important step, interpretation.

The central novelty of the model is that it regards attitude statements, which are the primitive terms of analysis in most studies of public opinion, as the outcome of a process involving entities that are even more primitive -- considerations or issue frames. Despite this novelty, however, that model is capable of incorporating the principal substantive insights of earlier models. For example, Converse's "non-attitudes" thesis is that for many individuals, the selection of a survey response is essentially a random shot. Nothing in our model necessarily denies that view. There can easily exist some (or perhaps many) individuals who can muster competing points of view, but who attach little weight to any of them and who vacillate stochastically between them. Such people may not be consciously flipping mental coins when they choose a response to a closed-ended question, but it is entirely consistent with our model to say, as Converse might wish to, that such individuals might just as well be flipping mental coins for all the conviction that their responses carry.

We emphasize, however, that we presently know very little about the mental structures that underlie survey responses. In the present instance, we cannot tell the difference between individuals who are deeply torn between competing considerations, and people who glibly mention several different ideas simply because they have been invited to talk. This is clearly a matter for further investigation. Our main point here is only that our model of the survey response is readily compatible with the central substantive claims of Converse's belief systems argument.

A central point of the measurement model tradition -- that response instability is due to vague questions -- is also consistent with our model. The reason, after all, that individuals so often bring different frames to the interpretation of the same question is that most questions are open to multiple interpretations. (Alternatively, one might say that most issues are multidimensional.)

Most users of measurement models also claim that individuals have "true attitudes" which are essentially perfectly stable. If one takes the central tendency, or average, of the competing frames in an individual's mind to be his "true attitude," then our model is again consistent with an important claim of the measurement model tradition. For as Feldman (1985) has pointed out, the

average of numerous considerations would be statistically indistinguishable from the "true attitudes" that are estimated by the measurement models.

In acknowledging points of contact with alternative models, however, we do not wish to obscure the central difference. That difference is that our model provides a much fuller specification of the process that leads to response instability, and in so doing, provides a theoretical opening for studying it.

Consider the matter of "attitude crystallization." Within our framework, one might take a crystallized attitude to be one that is based on a homogeneous set of considerations. Such homogeneity could be measured independently of response stability and then used to predict stability. Homogeneity of considerations could also perhaps replace response stability as the dependent variable in empirical investigations of this topic. In a similar way, one could study error variance as a variable function of the considerations in individuals' minds -- though, if one were successful in this undertaking, one would no longer want to call it error variance.

CONCLUSION

Humility comes quickly to those trying to understand the inner workings of the political mind. Not only is the subject itself unusually treacherous, but there has been a tendency to use experimental manipulations as a research tool, and experiments are relentlessly efficient debunkers of bad ideas. We speak of this from experience. The 1987 Pilot carried two experiments which, we had hoped, would yield useful insight into the nature of survey responses. Both failed.³⁴ Such unhappy outcomes are quite common in this field. Hence, until the proposed model is tested over a wider range of issues and conditions, it is prudent to regard it as tentative. In particular, our argument for the dominant consideration model is presently very far from being conclusive.

³⁴ One was an attempt to manipulate expressed approval of President Reagan by placing the approval item in either a favorable or unfavorable context. Our initial examination of the data suggests that the results came out in the expected direction, but very weakly. The other was an attempt to induce greater response stability for a random half of the sample the "stop and think" probes, as described earlier.

We nevertheless consider the model an extremely promising one. Its basic elements have appeared in the work of several different researchers, including that of individuals (Hochschild, Kelley) not concerned with the problem under investigation here. But more importantly, the model holds the promise of transforming a whole series of "methodological artifacts" into sources of insight into the nature of political attitudes (Tourangeau and Rasinski, 1986; Zaller, 1984). Given the magnitude of these supposed artifacts, this promise makes the model worthy of continued attention.

- Achen, Christopher (1975). Mass political attitudes and the survey response. American Political Science Review, 69, pp. 1218-1231.
- Bishop, George, R. W. Oldendick and A. J. Tuchfarber (1978). Effects of question wording and format on attitude consistency. Public Opinion Quarterly, 42, pp. 81-92.
- Bishop, George, R. W. Oldendick and A. J. Tuchfarber (1982). Political information processing: Question order and context effects. Political Behavior, 4, pp. 177-200.
- Bishop, George, R. W. Oldendick and A. J. Tuchfarber (1984). Interest in political campaigns: "What must my interest in politics be if I just told you I don't know. Public Opinion Quarterly, 48, pp. 510-519.
- Bishop, George, R. W. Oldendick and A. J. Tuchfarber (1985). The importance of replicating a failure: order effects on abortion items. Public Opinion Quarterly, 49, pp. 105-114.
- Converse, Philip (1964). The nature of belief systems in mass publics. In D. Apter (ed.), Ideology and Discontent. Free Press: New York, pp. 206-261.
- Converse, Philip E., and Gregory Markus (1979). Plus ca change...: the new CPS panel study. American Political Science Review, 73, pp. 32-49.
- Dean, Gillian and Thomas Moran (1977). Measuring mass political attitudes: change and unreliability. Political Methodology, 4, pp. 383-401.
- Erikson, Robert (1979). The SRC panel data and mass political attitudes. British Journal of Political Science, 9, pp. 16-49.
- Ericsson and Simon (1984). Protocol Analysis.
- Feldman, Stanley (1985). Reliability and stability of policy Positions: Evidence from a five-wave panel. Paper delivered at the annual meetings of the American Political Science Association, New Orleans, La.
- Hochschild, Jennifer (1981). What's Fair? Americans' Attitudes Toward Distributive Justice. Cambridge: Harvard University Press.
- Graeber, Doris (1985). Processing the News. New York: Longman.
- Kelley, Stanley, Jr. (1983). Interpreting Elections. Princeton: Princeton University Press.
- Loftus, E.F., S.E. Feinberg, and J. M. Tanur (1985). Cognitive psychology meets the national survey. American Psychologist, 40, pp. 175-180.
- Schuman, Howard and Stanley Presser (1981). Questions and Answers in Attitude Surveys. New York: Academic Press.
- Schuman, Howard and Jacqueline Scott (1987). Problems in the use of survey questions to measure public opinion. Science, 236, pp. 957-959.

Taylor, Shelley E., and Susan Fiske (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (ed.), *Advances in Social Psychology*, New York: Academic Press, pp. 249-288.

Tourangeau, Roger and Kenneth Rasinski (1986). Cognitive processes underlying context effects in attitude measurement. Unpublished manuscript, NORC, University of Chicago.

Tversky, Amos and Daniel Kahneman (1982). The framing of decisions and the psychology of choice. In Robin Hogarth (ed.), Question Framing and Response Consistency. San Francisco: Josey-Bass, pp. 3-20.

Zaller, John R. (1984). Toward a theory of the survey response. Paper delivered at the annual meetings of the American Political Science Association, Washington, D. C.

Zaller, John R. (1987). Selective attrition in panel studies. Memo to the Board of Overseers of the National Election Studies, Ann Arbor, Michigan.