RESPONSE-ORDER EFFECTS IN PUBLIC OPINION SURVEYS: THE PLAUSIBILITY OF RIVAL HYPOTHESES[1]

George F. Bishop, University of Cincinnati and Andrew E. Smith, University of Wisconsin-Milwaukee
George F. Bishop, PO Box 210375, University of Cincinnati, Cincinnati, OH 45221-0375

Key Words: Survey, Response-order Effects

As with so many other aspects of survey questions, the order in which the response alternatives are presented can significantly alter the results and conclusions of public opinion polls. Beginning with Schuman and Presser's (1981) revitalization of this line of research, a number of investigators have reported a variety of split-ballot experiments demonstrating how the order of the response alternatives produces a recency effect (choosing the last mentioned alternative), a primacy effect (choosing one of the first mentioned alternatives), or various kinds of judgmental contrast effects (see Knäuper, 1995; Krosnick, 1992; Krosnick and Alwin, 1987; Krosnick, *et al.*, 1996; McClendon, 1986; Mingay and Greenwell, 1989; Moore and Newport, 1996; Schwarz, *et al.*, 1992; Smith and Bishop, 1996; Sudman, *et al.*, 1996).

Such effects, particularly recency effects, as Moore and Newport (1996) have shown with their split-ballot experiments on questions about public policy issues in Gallup's telephone polls, can occur much more frequently, and with greater magnitude, than has heretofore been recognized. So the problem has become a central concern for polling practitioners as well as survey methodologists.

Rival Hypotheses

The Cognitive Elaboration Model

Explanations rooted in cognitive models of the question-and-answer process have, in recent years, provided a plausible account of how response-order, and other question effects, surface in survey data (see Hippler, *et al.*, 1987; Krosnick, 1991, 1992; Krosnick and Alwin, 1987; Krosnick *et al.*, 1996; see Sudman, *et al.*, 1996 for a state-of-the-art survey of the literature). One of the more compelling theoretical accounts of the response-order phenomenon has come from the work by Norbert Schwarz and his associates (Schwarz *et al.*, 1992, 1994; Sudman *et al.*, 1996; see also Knäuper, 1995): what is now known as the cognitive elaboration model. In the most current statement of the model (see Sudman *et al.*, 1996, Ch. 6) the mediating process that theoretically underlies response-order effects is the opportunity the respondent has to think about (elaborate on) the implications of the content of the response alternatives presented to him or her. This opportunity depends on a complex interaction of three things:

1. the serial position in which the response alternative is presented -- i.e., at or near the beginning, the middle, or the end of a list of alternatives;
2. the mode in which the response categories are presented: in an auditory format where they are read aloud to the respondents (with no visual aid); in a visual format such as in a self-administered questionnaire or in a face-to-face interview where the alternatives are presented on a "show card"; or in a combined auditory and visual format in which, for example, the response alternatives are read aloud to the respondent while being presented simultaneously on a show card; and
3. the plausibility of the response alternative, with a plausible alternative being one that "elicits agreeing thoughts when the respondent thinks about it and 'implausible' if it elicits disagreeing thoughts..." (Sudman *et al.*, 1996, pp. 140-41).

With other relevant factors held constant (memory limitations, complexity and extremity of response alternatives, and the respondent's ability and motivation), the elaboration model makes the following predictions:

1. In a visual presentation format, the model predicts largely primacy effects because respondents have a greater opportunity to think about (elaborate on) the response alternatives displayed at or near the beginning of the list than those listed later. But if the response alternative(s) presented early in the list is "implausible" (because it elicits disagreeing thoughts), the model predicts the opposite: a recency effect (see Sudman, *et al.*, 1996, p. 41).
2. In an auditory presentation format, the model predicts the reverse: mostly recency effects since respondents have a better opportunity to think about an alternative presented late in the list, especially if it is read aloud as the last choice, than those offered earlier in the list. But if the alternative(s) presented at the end is "implausible" (elicits disagreeing thoughts) the model predicts a primacy effect.

Other factors, such as memory limitations, judgmen-

[1]The authors wish to thank Ken Rasinski of NORC for his comments and recommendations on an earlier version of this paper.

tal contrast effects, and "confirmatory bias" can complicate these predictions, but the bulk of the evidence from a secondary analysis of split-ballot experiments carried out by the Allensbach Institute in Germany (Schwarz *et al.* 1992) would appear to support the cognitive elaboration model described by Sudman *et al.* (1996). A further replication by Schwarz *et al.* (1992), which varied Payne's (1951) classic experiment with the "oil supply" questions, also supports the hypothesis that response-order effects can be reduced or even eliminated if respondents are given an opportunity to think about the same subject in answering preceding questions as this increases "the temporary cognitive accessibility of relevant thoughts". And while there may not be sufficiently controlled experiments to test the elaboration model in its entirety, the evidence to date would seem to be essentially consistent with it.

Satisficing Theory

An equally compelling cognitive explanation of response-order effects has emerged from the work of Jon Krosnick and his associates (Krosnick, 1992; Krosnick and Alwin, 1987; Krosnick *et al.*, 1996): satisficing theory. Based on Simon's (1957) satisficing explanation of how people make economic decisions, this theory assumes that most respondents answer survey questions by choosing the first satisfactory or acceptable response alternative that is offered to them rather than take the time to select an optimal answer. The tendency to satisfice in this manner depends upon three things (Krosnick *et al.*, 1996):

1. the difficulty of the question-answering task due, for example, to the complexity of the topic and its familiarity or unfamiliarity to the respondent.
2. the respondent's ability to retrieve, process, and integrate information from memory; and
3. the respondent's motivation, in particular his or her "need for cognition"--the extent to which they like to think--as well as the personal relevance of the topic.

In the case of response-order, the theory predicts a form of "weak satisficing" that leads respondents to choose the first acceptable alternative that is presented to them in a closed-ended question, depending upon the mode in which the data are collected. If the response alternatives are read aloud without a visual aid (auditory mode) satisficing theory--like the cognitive elaboration model--predicts largely recency effects because the respondent can process more deeply (do more thinking about) the final alternative(s) read to them. But if the choices are presented in a purely visual format, as with "show cards", satisficing theory--again, much like the cognitive elaboration model--predicts primacy effects because respondents are able to think more about, and therefore process more deeply, the alternatives presented early in the list (see Krosnick, 1991; Krosnick and Alwin, 1987 for a more detailed account). The initial evidence from various experiments and meta-analyses by Krosnick and his associates has generally confirmed satisficing theory's predictions on response-order effects, demonstrating (1) that a primacy effect is most likely to occur when response alternatives are presented on a "show card" in a face-to-face interview (Krosnick and Alwin, 1987) and (2) that both primacy and recency effects tend to be stronger among respondents who are less educated, less cognitively sophisticated, and less cognitively skilled (see Krosnick, 1992; Krosnick *et al.*, 1996).

One noteworthy anomaly in these studies, however, was Krosnick's discovery of a significant primacy effect for a closed-ended questions about desirable qualities in a child when the five response alternatives were presented by telephone, whereas satisficing theory, as well as the cognitive elaboration model, would predict a recency effect in auditory mode (Krosnick, 1992, pp. 207-210). Such anomalies, as we will see, pose a significant challenge to the cognitive paradigm.

Testing the Generality of Rival Explanations

At this stage of theoretical development, it would clearly be useful to assess the plausibility, parsimony, and general explanatory power of these explanations of response-order effects. The availability of a sizable number of split-ballot experiments (over 200) on response-order conducted originally by the Gallup Organization provides a critical opportunity to test their generalizability across time and across a great variety of topics. In the analyses which follow, we will demonstrate that there are significant anomalies in the Gallup studies that cannot be accounted for by existing cognitive explanations and that there is therefore a need for a more parsimonious account of the response-order phenomenon. In the concluding discussion we propose still another plausible rival hypothesis.

The Gallup Split-Ballot Experiments

Since the late 1930s, Gallup conducted thousands of split-ballot experiments on nearly every aspect of question form, wording, and context, not to mention numerous "secret ballot" experiments on voting intentions (Bishop and Smith, 1991). The vast majority of experiments were done during the 1930s and 1940s. Of these, over 200 varied the order of the response alternatives on an extensive set of topics, ranging from simple trial heat questions on voting preferences to complex issues of domestic and foreign policy (see table 1 below). We excluded 41 experiments from our analysis because changes in response order were confounded with changes in ques-

tions wording, question order, or both. For this paper, we analyzed 213 unconfounded experiments. In most of these experiments (181), questions were read aloud to respondents without any visual aid (auditory mode). In the remaining experiments (32) the respondents were given a show card on which the response alternatives were presented (visual mode).[2] The typical sample size in most of these experiments was 2000-3000, so the likelihood of detecting statistically significant differences in these investigations is relatively high, even though the differences may be fairly small in magnitude. Let us now look at the results from these long neglected datasets and what they tell us about the generality of cognitive explanations of response-order effects in public opinion surveys.

Findings

A little over a fifth (22.5%) of the experiments generated a statistically significant response-order effect at the .05 level (see table 1). The average effects size for these studies was, however, relatively small in magnitude--typically less than 5%. Notice too that the average effect size varied with the topic.[3] For seemingly more familiar or less abstract subjects, such as presidential candidate preferences in "trial heat" questions and political party preferences of various types, the order of the response alternatives produced, on average, a fairly small effect. For more complex topics, the average effect size tended to be somewhat larger, thus providing some support for the conventional question complexity hypothesis, that response-order effects are a consequence of asking long, difficult, or complex questions (Payne, 1951; Rugg and Cantril, 1944; see also McClendon, 1986; Mingay and Greenwell, 1989; Moore and Newport, 1996).

Table 1: Average Effect Size and Number of Significant Response-Order Effects by Topic in Gallup Polls, 1936-1960

Topic	No. of Exper-iments	Avg. Effect Size	# Statist. Signif.
Cand. Trial Heats	51	2.8%	12
Pol. Party Preferences			
Party Identification	23	1.9	2
Party Vote Trial Heats	20	2.0	2
Party Like Win Elec.	14	2.1	2
Party Best for Groups	12	1.2	0
Party Best for Self and Reference Group	8	1.9	0
Party Best for Dealing w/Problem or Issue	9	2.9	2
Ideological Preferences	6	2.6	0
Left/Right Differences	3	6.2	1
Opinions on WWII Issues	38	2.9	13
Opinions on Other Domes. and Foreign Topics	12	4.4	6
Opinions on Topics of General Interest	17	3.5	8
Total	213		48

Recency and Primacy Effects by Mode

Table 2 shows the overall distribution of recency and primacy effects by presentation mode, when the difference by question form is simply classified by the direction of the difference (recency or primacy) without regard to statistical significance. Both the cognitive elaboration model and satisficing theory predict mostly recency effects when response alternatives are presented in auditory mode. But the results from 181 Gallup experiments conducted in this mode indicate that response-order effects in the recency direction were only slightly more likely to occur than primacy effects--in fact, not much different than would be expected by chance. Furthermore, the average effect size was identical for both primacy and recency effects (2.6), still another indication of chance fluctuation.

[2]Because the documentation of these early experiments is at best sketchy, it is difficult to determine with any degree of certainty whether experiments involving a "show card" were presented purely in visual mode. It is possible that in some instances, the interviewer read the question to the respondent as well as presenting a show card.

[3]The average effect size is the weighted average of the individual effect sizes calculated independently for each experiment. For an individual experiment, the effect size represents the largest between-form difference for one of the offered responses. Effect sizes are weighted by the number of cases in the experiment so that an experiment with 3000 cases, for example, would contribute proportionally more than an experiment with only 500 cases.

Table 2: Distribution of Recency and Primacy Effects by Presentation Mode

	Experiments #	Experiments %	Average Effect Size
Auditory Mode			
Recency effects	94	51.9	2.6
Primacy effects	87	48.1	2.6
Total	181	100.0	
Visual Mode			
Recency effects	18	56.3	3.5
Primacy effects	14	43.8	2.9
Total	32	100.1	

The results in visual mode pose an even greater problem for the cognitive models, which predict largely primacy effects when response alternatives are presented on show cards as was done in the 32 Gallup experiments summarized in table 2. A majority (56.3%) of the effects in this mode were recency effects rather than primacy effects. The average effect size in visual mode was also somewhat larger when the effect was in the recency, than in the primacy, direction. Though neither of the cognitive models makes explicit predictions about effect size by direction, these data nonetheless present an explanatory challenge for each of them.

If we look at just those experiments that were statistically significant (.05 level), the results vary somewhat, but still pose a substantial explanatory problem for cognitive models (table 3). Recency effects, as predicted, occurred more frequently than primacy effects in auditory mode (59% vs. 41%), but the latter are far too numerous to be explained away, ad hoc. The difference in average effect size for both recency and primacy effects in auditory mode was also negligible, again suggesting random fluctuation, especially when we consider the greatly reduced number of experiments in table 3 (39) as compared to those in table 2 (181).

Table 3: Distribution of Statistically Significant Recency and Primacy Effects by Presentation Mode

	Experiments #	Experiments %	Average Effect Size
Auditory Mode			
Recency effects	23	59.0	5.6
Primacy effects	16	41.0	5.2
Total	39	100.0	
Visual Mode			
Recency effects	6	75.0	5.7
Primacy effects	2	25.0	7.4
Total	8	100.0	

In the visual mode, the results look equally anomalous. Six of the eight experiments that generated a statistically significant difference yielded a recency effect, contrary to what would be generally predicted by the cognitive models. The average effect size, however, appeared to be somewhat greater for the two primacy effects than for the six recency effects. But given the small number of experiments (8) involved, chance fluctuation provides as compelling an explanation for the results in table 3 as it does for those in table 2. Recency and primacy effects seem nearly equally likely to occur regardless of presentation mode and with about the same average effect size. And if there is any departure from chance variation, it does not appear to be especially predictable from current cognitive models of response-order effects.

Can the Anomalies Be Explained?

Proponents of the cognitive elaboration model might argue that the anomalous results in tables 2 and 3 arise because the response alternatives in these experiments are somehow "implausible". In auditory mode, the model predicts mostly recency effects, but if the response alternatives are "implausible"--i.e.,they elicit disagreeing thoughts--it predicts the opposite: a primacy effect. And vice versa, in visual mode, the model predicts largely primacy effects, unless response alternatives are "implausible", in which case it predicts the opposite: a recency effect (see Sudman *et al.*, 1996, pp. 140-41). Such an argument can always be made post hoc, especially for archival data such as those presented here, in which there are no measures of whether respondents found response alternatives implausible because of "disagreeing thoughts".

How do our results speak to such an "implausibility" argument? Looking at just the response-order effects that were statistically significant (table 3), the implausibility argument has to be invoked 16 times to account for the unpredicted primacy effects in auditory mode and six times to explain away the unexpected recency effects in visual mode. Could all of these anomalies be the result of implausible response alternatives creating disagreeing thoughts?

It is likewise difficult to see how the satisficing theory of response-order effects can be sustained as it makes the same predictions about recency and primacy effects in auditory and visual mode and, unlike the elaboration model's implausibility proposition, provides no obvious way to account for the anomalies in tables 2 and 3.

An Alternative Explanation

At this point, neither of the cognitive models appear

to provide a plausible explanation for the various response-order effects in the Gallup experiments reported here, though each can point to some supporting evidence. The sheer number of experiments in the Gallup data which do not fit the predictions made by the cognitive elaboration model or satisficing theory lead us to look elsewhere for the solution to the riddle of response-order effects.

An important clue to unlocking the response-order phenomenon comes from a closer look at the experiments that produced the largest effects in the theoretically predicted direction, as well as those generating the most notable anomalies. Some of the largest examples of recency effects in auditory mode, for example, come from Moore and Newport's (1996) experiments with public policy issues. Nearly all of the questions that generated a sizable response-order effect in their study took a form much like the following example:

> "Which of the following reasons better describes why you disapprove of the agreement to reopen the government: because you wish the shutdown had continued until a balanced budget agreement was reached, **or** because the agreement did not go far enough in reopening the government?" [emphasis ours]

This particular question, administered by telephone (auditory mode) produced a highly significant and sizable recency effect: a 16% difference. As Moore and Newport (1996) observed, all of the significant effects they discovered were recency effects involving the use of lengthy and complex response alternatives--separated at the end, we would add, by the telling conjunction, **or**.

Similarly, McClendon (1986) found that all seven of the questions he used in a telephone survey resulted in recency effects, all of which took the form of the following example (which produced a 15% difference):

> "Should government have some responsibility for regulating health and safety conditions in the workplace, **or**, should private companies, unions, and workers be entirely responsible for regulating health and safety conditions in the workplace?" [emphasis ours]

Schuman and Presser's (1981) experiments with questions about the "oil companies" and "oil supply", each of which produced significant recency effects via the phone, also included as part of the wording either the conjunction, **or**, or the even more suggestive phrase about which response alternative is "nearly right", as seen below.

Oil Companies

> "Which of these statements comes closest to your own idea of how gasoline and oil prices are decided: each company sets its own prices to meet the competition, **or** the oil companies get together and set prices for their products." [emphasis ours]

Oil Supply

> "Some people say that we will still have plenty of oil 25 years from now. Others say that at the rate we are using our oil, it will all be used up in about 15 years. Which of these ideas would you guess is most nearly right?"

Both of these questions consist of lengthy, complex response alternatives, but sizable recency effects in auditory mode occurred, as we know from Schuman and Presser's research, even for short and simple questions on a topic that nearly everyone is familiar with, if not opinionated about, namely divorce:

> "Should divorce in this country be easier to obtain, more difficult to obtain, or stay as it is now?"

Notice that this item, which generated relatively large response-order effects (10-12% in Schuman and Presser's studies), contains the conjunction, **or**, which we will argue provides a cue to those respondents who are uncertain about how they should answer the question, that the choice which follows is the socially preferred, correct, or acceptable one. How else can this well-established response-order effect be explained? Memory limitations would hardly come into play on such a simple item. And with such short and simple response alternatives, all respondents would have plenty of time to think about each one and its implications (the elaboration model), or process each one more "deeply" (satisficing theory).

Still another clue can be found in Knäuper's (1995) reanalysis of Schuman and Presser's (1981) experiments with the divorce item, showing that the order effect is significantly and substantially larger among older, than younger, respondents. The reason for this, we would argue, has little or nothing to do with greater memory limitations among older respondents on this simple question, but rather with the presumably greater sensitivity of such respondents to giving a socially correct or acceptable answer to the socially sensitive question of divorce. For those respondents who are ambivalent about how they should answer this question, the conjunction, **or**, provides a cue, as it often does in everyday conversations, that the choice or alternative which follows is the preferred, correct, or acceptable one. And this, we think, is the basic principle that underlies response-order effects in general: those who are ambivalent or uncertain about how to answer a question (because of its sensitivity, difficulty, or complexity) will choose what seems to be the first socially correct or acceptable alternative. It is this subgroup, we believe, that accounts for the emergence of most response-order effects in public opinion surveys.

Consider, finally, the example from the present

study of the unexpected, sizable primacy effect in auditory mode on the question about the future of atomic energy (table 4). This brief item does not include any suggestive cues like the conjunction, **or**, but the question is stated in a way that makes it easy for those respondents who are uncertain about how to answer to just say "yes" to the somewhat leading proposition presented to them. "Going along" with the first alternative (good or harm) becomes the socially correct or acceptable solution to answering the question.

Table 4: Example of Unexpected Primacy Effect in Auditory Presentation Mode

Form I: "Do you think that, in the long run, atomic energy will do more good than harm?"
Form II: "Do you think that, in the long run, atomic energy will do more harm than good?"

	Form I	Form II	% Diff.
More good	68.8%	56.5%	12.1%
More harm	31.4	43.5	
	100.0	100.0	
(N=)	(1082)	(1035)	

X^2=32.94, df=1, P<.0001

Our alternative explanation of response-order effects is admittedly speculative and ad hoc. But we find that none of the rival explanations available can handle the anomalies identified here. Not, that is, without a lot of elaborate theorizing about cognitive processes in the black box, "implausibility", and the like. Far more parsimonious, we would argue, to propose that respondents who are uncertain about how to answer a survey question (for whatever reason) will, using cues contained in the wording and structure of the question (e.g., **or**), choose the first socially correct, or acceptable, response alternative presented to them. This is the principle that we believe underlies response-order effects in general, one that can be readily tested and falsified.

References

Bishop, G., and A. Smith. 1991. "Split-Ballot Experiments in the Gallup Poll: 1938 - 1988." The Public Perspective 2:25-27.

Hippler, H., N. Schwarz, and S. Sudman. 1987. Social Information Processing and Survey Methodology. New York: Springer-Verlag.

Knäuper, B. 1995. "Response Effects in Older Age." Paper presented at the annual conference of the Midwest Association for Public Opinion Research, Chicago, Illinois, October.

Krosnick, J. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." Applied Cognitive Psychology 5:213-236.

Krosnick, J. 1992. "The Impact of Cognitive Sophistication and Attitude Importance on Response-Order and Question-Order Effects," in N. Schwarz and S. Sudman, eds., Context Effects in Social and Psychological Research. New York: Springer-Verlag, pp. 203-218.

Krosnick, J. and D. Alwin. 1987. "An Evaluation of Cognitive Theory of Response-Order Effects in Survey Measurement." Public Opinion Quarterly 51:201-219.

Krosnick, J. S. Narayan, and W. Smith. 1996. "Satisficing in Surveys: Initial Evidence," in M. Braverman and J. Slater eds., Advances in Survey Research. San Francisco: Jossey-Bass, No. 70, Summer, pp. 29-44.

McClendon, M. 1986. "Response-Order Effects for Dichotomous Questions." Social Science Quarterly 67:205-211.

Mingay, D. and M. Greenwell. 1989. "Memory Bias and Response-Order Effects." Journal of Official Statistics 5:253-63.

Moore, D. and F. Newport. 1996. "Public Policy Questions and Response Order: Prevalence of the Recency Effect." Paper presented at the annual meeting of the American Association for Public Opinion Research, May.

Payne, S. 1951. The Art of Asking Questions. Princeton: Princeton University Press.

Rugg, D. and H. Cantril. 1944. "The Wording of Questions," in H. Rugg ed., Gauging Public Opinion. Princeton: Princeton University Press, pp.23-50.

Schuman, H. and S. Presser. 1981. Question and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Schwarz, N., H. Hippler, E. Noelle-Neuman. 1992. "A Cognitive Model of Response-Order Effects in Survey Measurement," in N. Schwarz and S. Sudman eds., Context Effects in Social and Psychological Measurement. New York: Springer-Verlag.

Schwarz, N., H. Hippler, E. Noelle-Neuman. 1994. "Retrospective Reports: The Impact of Response-Alternatives," in N. Schwarz and S. Sudman eds., Autobiographical Memory and the Validity of Retrospective Reports. New York: Springer-Verlag.

Simon, H. 1957. Models of Man. New York: Wiley.

Sudman, S., N. Bradburn, and N. Schwarz. 1996. Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.