

*Two split-ballot experiments on attitude questions—one on inclusion or exclusion of “don’t know” options and one on agree/disagree versus forced-choice format—were included in the General Social Survey in 1974 and replicated in 1982. Response effects occurred in each experiment in 1974 and were generally replicated in 1982, but the effects do not interact with time. Despite the significant response effects, both forms for each experiment yield similar conclusions about change over time, as long as question wording is kept constant. However, a borderline interaction with education in the 1974 data for the agree/disagree versus forced-choice experiment replicates more clearly in 1982, indicating that the two forms yield different conclusions about the relation of the question content to education.*

## Response Effects Over Time:

### Two Experiments

HOWARD SCHUMAN

JACQUELINE SCOTT

*University of Michigan*

### INTRODUCTION

Based on many split-ballot experiments, survey research specialists know that what might ordinarily be considered nonsubstantive variations in questions can often shift univariate distributions (“marginals”) substantially (Cantril, 1944).<sup>1</sup> Less frequent are statistically significant differences in associations between two such nonsubstantive variations of the “same” question and standard background variables (e.g., education), but some interactions of this kind have also been documented (Schuman and Presser, 1981). We know least of all, however, about interactions of nonsubstantive question variations with time when one attempts to study changes in attitudes, values, or beliefs. Schuman and Presser (1981) were able to report only two cases where adequate split-ballot experiments were repeated over a lengthy

---

AUTHORS’ NOTE: *The research reported here was supported by a grant from the National Science Foundation (SES-8713633).*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 17 No. 4, May 1989 398-408  
© 1989 Sage Publications, Inc.

time interval. In one case, a wording variation that changed univariate distributions substantially did not reveal any sign of an interaction with time over a 34-year period because essentially the same response effect occurred at each time point (pp. 276-280). In the other case, however, an effect due to question order or context did interact with the 32-year trend, and an investigator using one question order would have drawn different conclusions about change than an investigator using the other question order (pp. 28-31).

The two examples reported by Schuman and Presser are of some theoretical interest, but neither involved survey items that are used today as important measures of social change. In 1982, however, the General Social Survey (GSS) repeated two split-ballot experiments that had been carried out originally in the GSS in 1974 with standard items used to assess trends: In one case, a question measured attitudes toward punishment of criminals by the courts (the COURTS question); in the other case, a question measured attitudes toward participation in politics by women (the FEPO question).<sup>2</sup> Both of the 1974 experiments had revealed significant univariate response effects, and each had also shown a less certain interaction with education. Here the interest is in whether these earlier findings occurred also in the 1982 survey, and even more important, whether the effects due to variations in question wording interacted with time in a way that would lead to different conclusions about social change depending on the question version used.

It is not difficult to imagine why such interactions with time could occur. For example, one of the most common variations in question wording is inclusion or exclusion of an explicit "don't know" (DK) alternative, inclusion to legitimate DK choices, and exclusion to discourage such choices (this variation is the basis of the COURTS experiment; see Table 1). It seems likely that the effect of omitting such a DK filter alternative would show little or no change over time, because the respondents are being implicitly told that DK responses are not invited regardless of subject matter. But when a DK filter is included, the proportion of respondents who elect to say "don't know" might well vary, depending on the salience of a particular issue at a specific point in time. For example, an issue that might have been obscure in 1974 might have become much more prominent a decade

later, with a consequent decrease in DK responses. Moreover, if those who had formerly answered DK do not distribute themselves into substantive categories in the same way as do other respondents, this would also lead to a change over time in the substantive responses to the filtered version of the question.

Much the same reasoning can apply to other nonsubstantive variations in survey items, and thus the possibility of interactions with time of nonsubstantive variations in question wording constitutes an important methodological issue that confronts those who use surveys to study social change. Empirical evidence is much needed.

### *THE EXPERIMENTS*

The 1974 GSS employed a randomized split-ballot to test nonsubstantive alterations in two regular GSS questions:

#### *COURTS*

The standard COURTS question to half the sample read as follows: "In general, do you think the courts in this area deal too harshly or not harshly enough with criminals?"

The experiment consisted of paralleling the standard question with a variation asked of the other half of the sample with a "don't know" filter added: "In general, do you think the courts in this area deal too harshly or not harshly enough with criminals, or don't you think you have enough information about the courts to say?"

#### *FEPOL*

The standard FEPOL question asks respondents whether they agree or disagree with the following statement: "Most men are better suited emotionally for politics than are most women."

The experiment consisted of paralleling the standard question with a new version that was written in balanced forced-choice form that sought to avoid possible acquiescence bias: "Would you say that most men are better suited emotionally for politics than are most women,

that men and women are equally suited, or that women are better suited than men in this area?"

Exactly the same two experiments were repeated in the 1982 GSS, and the analysis of change over time in these differences in question form (response by question form by time) constitutes the main focus of this article.

### *THE 1974 RESULTS*

As reported in Schuman and Presser (1981:120, 140) and included here in Table 1, the 1974 results showed the following.

"Don't know" responses were 22% higher when a DK-type filter alternative was added to the COURTS question, although the ratios of the three substantive responses to one another were not affected by the variation in question form. In addition, the DK response was more highly related to education on the standard form than on the filtered form ( $p < .05$ ), but because this was not true for most other similar experiments reported by Schuman and Presser (1981), the finding was regarded as of uncertain reliability.

For the FEPOI question, the percentage choosing the alternative represented by the GSS agree response ("Most men are better suited emotionally for politics than are most women") decreased significantly when it was part of a forced-choice question form (see Table 2). In addition, the agree/disagree version showed a stronger relation to education than the balanced forced-choice form, although the interaction was only borderline in significance ( $p < .10$ ); in three later telephone surveys, similar but nonsignificant ( $p > .10$ ) trends were also found.

### *COURTS*

Table 1 presents the results for each form of the COURTS question in both 1974 and 1982. There is a significant question-form effect on DK responses (DK versus substantive alternatives combined) in each year, but both forms show similar movement over time; therefore, there is no significant overall three-way interaction of time, question

TABLE 1  
Courts Question by Form and Year

Year	Standard		Filtered	
	1974	1982	1974	1982
<b>Response:</b>				
a. Too harshly	5.6*	2.6*	4.6*	4.4*
b. Not harshly enough	77.9	86.2	60.3	76.3
c. About right (volunteered)	9.7	7.5	6.1	4.9
d. DK/Not enough information to say	6.8	3.8	29.0	14.4
	100.0 (745)	100.0 (744)	100.0 (723)	100.0 (752)
<b>Models Tested:</b>				
		<u>df</u>	<u>L<sup>2</sup></u>	<u>p</u>
	{FR <sub>1</sub> }, {FY}, {R <sub>1</sub> Y}	1	0.89	ns
	{FR <sub>2</sub> }, {FY}, {R <sub>2</sub> Y}	2	2.88	ns
	{FR}, {FY}, {R <sub>Y</sub> }	3	3.78	ns
1974 only:	{F}, {R <sub>1</sub> }	1	130.89	<.001
	{F}, {R <sub>2</sub> }	2	1.14	ns
	{F}, {R}	3	132.03	<.001
1982 only:	{F}, {R <sub>1</sub> }	1	54.01	<.001
	{F}, {R <sub>2</sub> }	2	7.61	<.05
	{F}, {R}	3	61.62	<.001
1974, 1982 combined:	{F}, {R <sub>2</sub> }	2	5.86	<.10

NOTE: F = form (standard vs. filtered); R = response (a, b, c, d); R<sub>1</sub> = response (DK vs. substantive [d vs. {a, b, c}]); R<sub>2</sub> = response (a, b, c); and Y = year.

form, and DK responses. When the three substantive responses are analyzed separately from DK, again there is no three-way interaction of time by form by response.<sup>3</sup>

Unlike 1974, however, the 1982 data do show a significant difference by form for the three substantive responses ( $p < .05$ ). Because the interaction with time is not significant, the two time periods were collapsed and the combined sample (1974 plus 1982) yields a borderline effect of question form on substantive response:  $L^2 = 5.86$ ,  $df = 2$ , and  $p = .053$ . Further partitioning of the substantive subtable suggests that the main contrast contributing to the borderline effect is the choice of "too harshly" versus "about right." "Too harshly" is

TABLE 2  
**Women in Politics (FEPOL) Question by Form and Year<sup>a</sup>**

Year	Agree-Disagree		Forced-Choice	
	1974	1982	1974	1982
<b>Response:</b>				
Men better suited	47.0%	37.7%	33.1%	24.4%
Men & women equally suited	53.0	62.3	62.6	70.2
Women better suited	--	--	4.3	5.4
	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>
	(698)	(698)	(719)	(738)

**Models Tested:**

		df	L <sup>2</sup>	p
	{FR}, {RY}, {YF}	1	0.05	ns
1974 only:	{F}, {R}	1	22.12	<.001
1982 only:	{F}, {R}	1	22.88	<.001

NOTE: F = form (agree/disagree vs. forced-choice); R = response; and Y = year.

<sup>a</sup>Because the forced-choice version of the FEPOL question has three alternatives, structural zeros are included for the agree/disagree form, and the L<sup>2</sup> statistics reflect such tests. If "men and women equally suited" and "women better suited" are collapsed—as might be done in further analysis because of the small size of the "women better suited" category—the L<sup>2</sup> for the 1974 form-effect test is 28.6 (df = 1, ns), and for 1982, it is 29.8 (df = 1, p < .001); the L<sup>2</sup> for the three-way interaction is 0.08 (df = 1, ns). Thus, the results are essentially the same in the two approaches.

selected almost equally on the two forms, but "about right" is given less often on the filtered form. A quite plausible interpretation of this finding is that some of the people giving the "about right" response on the standard form (which omits an explicit DK alternative) do not have well crystallized views on the courts issue and therefore choose DK ("not enough information about the courts to say") when that alternative is explicitly offered on the filtered form. In sum, although there is no sign in Table 1 of an interaction with time, when the two years are combined there is some evidence that the choice of a particular

**TABLE 3**  
**Courts Question by Form and Education**

Education	Standard			Filtered		
	0-11	12	13+	0-11	12	13+
<u>GSS 1974</u>						
Response:						
a. Too harshly	6.6%	4.7%	5.6%	4.4%	3.4%	5.8%
b. Not harshly enough	81.5	78.0	73.6	58.6	69.4	52.9
c. About right	8.5	9.4	11.3	5.2	5.2	7.9
d. DK/Not enough information to say	3.5	7.9	9.5	31.7	22.0	33.3
	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>	<u>100.0</u>
	(250)	(254)	(231)	(249)	(232)	(260)

GSS 1982

Response:

a. Too harshly	3.8	1.3	2.6	6.3	2.9	4.4
b. Not harshly enough	85.9	89.6	83.1	71.2	82.5	73.9
c. About right	6.4	4.6	11.3	5.0	2.9	7.2
d. DK/Not enough information to say	3.8	4.6	3.0	17.6	11.8	14.5
	<u>(234)</u>	<u>(240)</u>	<u>(266)</u>	<u>(222)</u>	<u>(280)</u>	<u>(249)</u>

Models Tested	df	L <sup>2</sup>	$\Delta$ df	$\Delta L^2$	p
1974 only: {FR <sub>1</sub> }, {ER <sub>1</sub> }	3	14.57			
{FR <sub>1</sub> }, {ER <sub>1</sub> }, {FER <sub>1</sub> }	2	9.85	1	4.72	<.05
1982 only: {FR <sub>1</sub> }, {ER <sub>1</sub> }	3	3.13			
{FR <sub>1</sub> }, {ER <sub>1</sub> }, {FER <sub>1</sub> }	2	3.13	1	0.00	ns
1974 only: {FR <sub>2</sub> }, {ER <sub>2</sub> }	6	4.92			
{FR <sub>2</sub> }, {ER <sub>2</sub> }, {FER <sub>2</sub> }	4	4.16	2	0.76	ns
1982 only: {FR <sub>2</sub> }, {ER <sub>2</sub> }	6	15.17			
{FR <sub>2</sub> }, {ER <sub>2</sub> }, {FER <sub>2</sub> }	4	14.98	2	0.19	ns
{FR <sub>1</sub> E}, {FR <sub>1</sub> Y}, {FEY}	5	14.76			
{FR <sub>1</sub> E}, {FR <sub>1</sub> Y}, {FEY}, {FR <sub>1</sub> EY}	4	12.98	1	1.78	ns
{FR <sub>2</sub> E}, {FR <sub>2</sub> Y}, {FEY}	10	19.71			
{FR <sub>2</sub> E}, {FR <sub>2</sub> Y}, {FEY}, {FR <sub>2</sub> EY}	8	19.15	2	0.59	ns
{FRE}, {FRY}, {FEY}	18	38.56			
{FRE}, {FRY}, {FEY}	15	36.29	3	2.37	ns

NOTE: F = form (standard vs. filtered); R = response (a, b, c, d); R<sub>1</sub> = response (DK vs. substantive [d vs. {a, b, c}]); R<sub>2</sub> = response (a, b, c); Y = year; E = education treated as linear (0-11, 12, 13+ years); and  $\Delta$  = difference.

TABLE 4  
 Women in Politics (FEPOL) Question by Form and Education

Education	1974			1982		
	0-11	12	13+	0-11	12	13+
<u>Agree/Disagree</u>						
% Agree men better suited	56.6 (244)	44.5 (236)	38.7 (217)	54.5 (209)	36.7 (226)	24.3 (259)
<u>Forced Choice</u>						
% Choose men better suited	33.3 (243)	37.5 (232)	28.1 (242)	32.7 (211)	22.2 (275)	19.9 (251)
Odds Ratio	.38	.75	.62	.40	.49	.77

  

Models Tested <sup>a</sup>	df	L <sup>2</sup>	$\Delta$ df	$\Delta$ L <sup>2</sup>	p
1974 only: {FR}, {FE}	3	7.18			
{FR}, {FE}, {FER}	2	3.87	1	3.30	<.10
1982 only: {FR}, {FE}	3	5.97			
{FR}, {FE}, {FER}	2	1.38	1	4.59	<.05
{FER}, {FYR}, {FEY}	5	5.38			
{FER}, {FYR}, {FEY}, {FEYR}	4	5.25	1	0.13	ns

NOTE: F = form (agree/disagree vs. forced-choice); R = response ("men better suited" vs. "men and women equally suited"/"women better suited"); Y = year; E = education treated as linear (0-11, 12, 13+ years); and  $\Delta$  = difference.

<sup>a</sup>We combine "men and women equally suited" and "women better suited" for the models tested because the two alternatives appear only in the forced-choice form. If the two are distinguished and structural zeros are used in testing the interactions, there are only trivial differences for the figures reported here:  $L^2 = 3.14$  for 1974, 5.42 for 1982, and 0.12 for the four-way interaction.

substantive response is affected by question form, along with the larger effect of giving or not giving the DK alternative itself.

The 1974 data had also indicated an interaction involving DK, education, and response ( $p < .05$ ), but Schuman and Presser (1981:140) were skeptical of its reliability based on results from a number of experiments with other items. Table 3 shows that the interaction is not replicated in 1982, which suggests that what had appeared as an anomaly in 1974 was indeed the product of sampling error. There is also no four-way interaction of time (1974 versus 1982), education, form (standard versus filtered), and response, and the Table 3 data do not show any other significant interactions involving education.

*FEPOL*

Table 2 presents the results for each form of the FEPOL question at each time point. There is a highly reliable question-form effect on response, but both forms show much the same change over time; hence, there is no significant three-way interaction of time, form, and response. Thus there appears to be an acquiescence bias produced by the original GSS item, but it is constant over time and therefore does not affect the measurement of trends.

In addition, as is evident in Table 4, the relation of FEPOL to education has also increased over time on both forms, and there is thus no overall interaction of time, education, form, and response. In 1982, however, the relation of education to response is significantly stronger on the agree/disagree form than on the balanced forced-choice form, a similar but less reliable trend in the same direction having been found for the 1974 data. Thus, there is further evidence that acquiescence bias occurs especially among less educated respondents (note the odds ratios in Table 4); but because the degree of bias does not change between 1974 and 1982, there is no significant four-way effect involving time.

*CONCLUSIONS*

Both questions used in these experiments showed substantive changes over the eight-year period between the 1974 GSS and the 1982 GSS. Both questions also showed reliable response effects from variations in question form in each of the two years. Because the changes over time occurred to about the same degree on each form of each question, one would draw the same conclusion about the direction and approximate amount of substantive change regardless of which form of a question was used, *provided that question form was held constant in the comparison.*

This conclusion is consistent with one earlier experiment on response effects over time (Schuman and Presser, 1981:276-278), but it

cannot be guaranteed to occur for every type of response effect. The single documented exception at this point involves a context effect noted earlier, which clearly changes the meaning of the question; it seems possible that this is the only type of response effect likely to interact with time. The time period of eight years tested in the present analysis was not a long one, however, and it would be useful to repeat these two experiments in another decade. Because the present article provides further evidence of interactions involving question form and education, the possibility that trends over time can also be implicated remains viable.

Moreover, even if interactions that involve time are largely limited to context variations, one should note that such context effects are the hardest type of response effect to guard against. They can occur anywhere in a questionnaire, and are difficult to prevent because full prevention requires replication of the entire questionnaire verbatim. As surveys are increasingly used to measure social change, the problems dealt with here become more crucial to investigate.

## NOTES

1. Examples of supposedly nonsubstantive variations are: inclusion or exclusion in a question of a "don't know" response alternative; alterations in the ordering of questions; use of agree/disagree versus interrogative or forced-choice formats; and change in wording by substitutions with synonyms.

2. COURTS and FEPOL (that is, females in politics) are the mnemonics used in the GSS codebooks to identify these two questions (see Davis and Smith, 1988).

3. As these results suggest and as Table 1 shows, there is no overall interaction of time, form, and response, with the four response categories uncombined.

## REFERENCES

- CANTRIL, H. (ed.) (1944) *Gauging Public Opinion*. Princeton, NJ: Princeton Univ. Press.
- DAVIS, J. A. and T. W. SMITH (1988) *General Social Surveys, 1972-1988*. Chicago: National Opinion Research Center.
- SCHUMAN, H. and S. PRESSER (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.

*Howard Schuman is Professor of Sociology and Director of the Survey Research Center, University of Michigan. His research interests include the interplay of method and substance in surveys and collective memories about events and changes.*

*Jacqueline Scott is Visiting Assistant Professor of Sociology and 1989 Director of the Detroit Area Study, University of Michigan. Her research interests include parental stress, value conflicts in the abortion dispute, and collective memory.*