

Male and female telephone interviewers are compared on both administrative efficiency and data quality, using data from 24 replications of an attitudinal survey on personal and national economic prospects. The 40 male interviewers used over the two-year period are found to exhibit higher turnover rates and, because of that, lower response rates and higher training costs than the 80 female interviewers. However, there are no real differences on the total per minute interview costs by gender, in missing data rates, or on response distributions for factual questions. There does appear to be a systematic tendency for male interviewers to obtain more optimistic reports from respondents regarding their economic outlook. Multivariate models are constructed that attempt to explain these results and speculations are offered about causes of the impact of interviewer gender on response formation.

Gender Effects Among Telephone Interviewers in a Survey of Economic Attitudes

ROBERT M. GROVES
NANCY H. FULTZ
The University of Michigan

Past research in survey methodology has shown that personal characteristics of interviewers can affect both the likelihood of sample persons' cooperation with the survey and their responses to survey questions (e.g., Inderfurth, 1972; Hatchett and Schuman, 1975). Thus interviewer characteristics can be associated both with administrative efficiency and data quality. With a growing percentage of surveys being conducted over the telephone, those interviewer characteristics identifiable

AUTHORS' NOTE: An earlier version of this article was presented at the 1984 meetings of the American Association for Public Opinion Research. We appreciate the comments of Stanley Presser and the assistance of Richard Curtin in accessing the Survey of Consumer Attitudes data.

SOCIOLOGICAL METHODS & RESEARCH, Vol. 14 No. 1, August 1985 31-52
© 1985 Sage Publications, Inc.

by voice become increasingly important; of these, the interviewer's gender is the most obvious. This article examines interviewer gender as a personal characteristic that may affect the expense and quality of telephone survey data.

It is reasonable to suggest that male interviewers might obtain higher average refusal rates than female interviewers. Fear of crime has been used previously to explain nonresponse patterns in the United States (House and Wolf, 1978), and it is possible that the threat of an unknown male's voice would be perceived by respondents as greater than that of an unknown female's. Discussions with supervisors of telephone interviewing facilities indicate that male interviewers may also have a higher turnover rate than female interviewers, raising training costs for males. Correspondingly, the males' relative lack of experience might increase refusal rates even further.

In a study of telephone interviewers on the 1980 Farm Women's Survey, male interviewers were found to obtain significantly higher refusal rates from the sample of women when compared to female interviewers. In addition, the average length of an interview was significantly longer for the males (Nealon, 1983). An earlier summary of Census Bureau analyses of interviewer performance also noted that men completed fewer questionnaires per hour than did women on the 1964 Census of Agriculture (Inderfurth, 1972). These findings imply higher per unit costs for interviews conducted by males when interviewers are paid on an hourly basis.

Past research has produced mixed conclusions regarding the effect of interviewer gender on the quality of survey responses. Respondents to the Farm Women's Survey gave significantly different responses to both factual questions such as farm value and to attitudinal questions regarding satisfaction with farm programs, depending upon the gender of the interviewer. However, in two surveys on personal behavior, the nature of reports of sexual activities was not related to interviewer gender (Johnson and DeLamater, 1976). Female students gave more feminist responses to male than to female interviewers on questions of women's rights and roles (Landis et al., 1973), whereas male

respondents offered more conservative opinions on that subject to male interviewers (Ballou and Del Boca, 1980). Colombotos et al. (1968) found that male interviewers obtained higher scores on two psychiatric measures, from male as well as female respondents, although these differences were not significantly large. In another study, the percentage of respondents giving "sex habits" as a possible cause of mental disturbance was greater for female interviewers (Benney et al., 1956).

A review of the behavioral research literature by Rumenik et al. (1977) revealed that male rather than female experimenters appear to elicit better performance from adult subjects given verbal learning tasks. There was also some indication that clients provide more "information-seeking" responses to male counselors. However, experimenter gender effects were not clearly established in the data on clinical and psychological testing. In a decision-making simulation enacted with married couples, the wives were found to offer more problem-solving statements when the observer was a woman (Kenkel, 1961).

There are three major weaknesses with much of the research conducted on interviewer gender effects:

- (1) Completed interviews are often treated as independent observations, neglecting the fact that each interviewer generates a "cluster" of interviews. This oversight can result in severely biased estimates of standard errors of gender effect measures (Dijkstra, 1983).
- (2) The small numbers of men traditionally found on interviewing staffs threaten the generalizability of many conclusions in this area. The smaller the number of interviewers of each sex involved in a study, the greater the likelihood that any findings are the result of attributes peculiar to those few people.
- (3) Rarely do studies randomly assign sample cases to interviewers, and there is often a failure to control for additional variables that may explain an apparent interviewer gender effect.

This article attempts to overcome these obstacles by using estimators of standard errors for survey statistics that reflect the clustering among interviewers, by pooling data over time from an

ongoing survey to provide relatively large numbers of interviewers of each sex, and by using a design that randomized assignment of cases to interviewers.

STUDY DESIGN

Instead of reporting the results of a single experimental telephone survey examining gender effects, this article discusses 24 replications of the same survey, the Survey Research Center's Survey of Consumer Attitudes, over a two-year period (September 1981 to August 1983). The survey has a two-wave panel design whereby randomly selected respondents connected to sample telephone numbers (generated using a two-stage rejection rule RDD design; Waksberg, 1978) are reinterviewed six months after an initial interview. The survey is an ongoing one at the SRC centralized telephone facility; each month approximately 400 interviews are taken with first-time RDD respondents and 300 with second-time recontact respondents (see Steeh et al., 1983, for a full description of the surveys). Over the 24-month period 80 different female and 40 different male interviewers completed approximately 16,000 interviews, although the number of different respondents is only somewhat more than half that number. In order to reduce data processing costs, a one-third subsample of interviews taken by female interviewers was chosen, so that the analytic data set used contains about 7,300 interviews. The response rate among first-time respondents varied over the months between 64% and 77%, with a mean of 72%. These response rates include all confirmed household numbers in the denominator, including those not contacted but verified as working by local telephone business offices. There is no apparent trend toward higher or lower response rates over the 24-month period.

Our ability to examine gender effects using these data is enhanced by the fact that within interviewing shifts each interviewer is given a random subset of the active sample cases (those that have not yet yielded an interview or a final refusal). This randomization assures that the respondent groups contacted by interviewers in the same shift will be equivalent (in expectation) to

one another. It eliminates the possibility that at the time of initial assignment male interviewers are given particular types of respondents to contact and female interviewers other types (see Groves and Magilavy, 1983, for a description of this procedure).

The inferential model used in this article assumes that the group of interviewers used in the 24 months of data collection was selected from a large population of potential interviewers available to SRC. Inference is thus strictly limited to the SRC experience, but the results of the work are useful to others to the extent that their labor markets, training, supervision, and survey designs are similar to those used at SRC. Variability in gender effects over possible replications of this data collection effort are viewed to arise from different interviewers being selected from the pool. Thus measures of standard errors reflect the clustering of observations into interviewer assignments. To estimate standard errors for means and regression coefficients, jackknife estimators were used (Brillinger, 1964; Kish and Frankel, 1974). These estimators are based on pseudo-replicates of the data created by dropping successively one interviewer from among all interviewers. Estimates of interest are computed on each replicate, and the variability in those estimates across the replicates is measured. Such estimates of the standard errors have been found to have desirable properties in several empirical investigations.

When models are constructed on the data in the later sections of the article, the clustering of observations into interviewer assignments is explicitly reflected in parameter estimates through weighted least squares estimation, using a covariance matrix reflecting the clustering (see Landis et al., 1982, for a discussion of this technique). These matrices were estimated using a Taylor Series estimator, treating interviewer respondent groups as ultimate clusters.

GENDER DIFFERENCES IN OBTAINING RESPONDENT COOPERATION

Table 1 presents the initial refusal rates for male and female interviewers based on the first-time RDD sample. These rates are

TABLE 1
Initial Refusal and Overall Response Rate by Interviewer Gender

Tenure	Initial Refusal Rate ^a	Overall Response Rate ^b	Proportion of Observations
Male			
1-6 months	.257	.704	.605
7-12 months	.166	.715	.157
13 or more	.178	.816	.238
Overall	.224	.732	1.000
Female			
1-6 months	.227	.733	.332
7-12 months	.155	.809	.165
13 or more	.133	.835	.503
Overall	.168	.797	1.000

a. Refusal rate figures are the means of individual refusal rates achieved each month by interviewers on the Survey of Consumer Attitudes. The refusal rate is the ratio of initial refusals to the entire initial interviewer assignment.

b. Response rate figures are means of monthly individual response rates achieved by interviewers on the Survey of Consumer Attitudes. Response rate is defined as ratio of (Complete + Partial Interviews) to (Complete + Partial Interviews + Final Refusals + initial refusals not converted by others + other Noninterviews).

the averages (over months) of individual interviewer refusal rates obtained in specific months of the Surveys of Consumer Attitudes. Overall, male interviewers have refusal rates that are approximately five percentage points higher than those of female interviewers, consistent with the speculations presented earlier.

The gross comparison of refusal rates between the sexes glosses over, however, certain sources of variability in cooperation rates. Foremost among these are differences in performance related to experience as an interviewer. To address this, Table 1 also separates interviewers by their tenure on the survey. With this disaggregation it is clear that novice interviewers (1 to 6 months of tenure) have the highest initial refusal rates in both gender

TABLE 2
Demographic Characteristics of Respondents by Interviewer Gender

Characteristic of Respondent	Percentages				Number of Cases	
	Gender of Interviewer		Difference	Standard Error of Difference	Males	Females
	Male	Female				
Gender					2862	4345
Male	45.9%	42.6%	3.3%*	1.6		
Female	54.1	57.4	-3.3*	1.6		
Age					2848	4321
18-30	31.1	29.3	1.8	1.5		
31-40	25.5	22.5	3.0*	1.4		
41-60	27.5	27.2	0.3	1.4		
61 or over	15.9	20.9	-5.0*	1.2		
Race					2787	4222
White	88.9	88.2	0.7	1.0		
Nonwhite	11.1	11.8	-0.7	1.0		
Education					2740	4146
8 yrs. or less	5.8	7.5	-1.7	0.8		
9 - 11 years	10.3	11.1	-0.8	1.0		
High School	33.1	33.8	-0.7	1.5		
Some College	23.0	21.5	1.5	1.3		
Junior College	3.2	2.9	0.3	.6		
4 yrs College or more	24.5	23.2	1.3	1.4		
Employment Status					2686	3861
Working for Pay	63.4	57.3	6.1*	1.4		
Other	36.6	42.7	-6.1*	1.4		
Marital Status					2805	4244
Married	60.2	59.2	1.0	1.6		
Separated, Divorced	13.2	12.2	1.0	1.1		
Widowed	8.4	11.0	-2.6*	0.9		
Never Married	18.1	17.6	0.5	1.2		
Family Income					2735	4138
\$0 - \$9,999	16.3	19.2	-2.9*	1.2		
\$10,000 - \$14,999	12.1	11.5	0.6	1.0		
\$15,000 - \$19,999	11.7	11.6	0.1	1.0		
\$20,000 or more	59.9	57.7	2.2	1.6		

*Estimate of difference more than two times its standard error. Standard error estimates based on generalized jackknife replication estimates, reflecting the clustering of observations into interviewer assignments.

DIFFERENCES IN COST EFFICIENCY BETWEEN MALE AND FEMALE INTERVIEWERS

Design choices affecting survey errors are never made in the absence of cost constraints. For example, given fixed resources, the desire to maximize response rate (in hopes of minimizing nonresponse bias) may force a smaller sample size (and hence higher sampling variance) because of callback and refusal conversion costs. It is therefore useful to examine whether male and female interviewers are differentially cost effective. By "cost effi-

ciency" we mean the cost of obtaining one minute of interview data. Thus cost efficiency has two components: the rate per hour of obtaining interviews, and the salary and training costs incurred per minute employed.

INTERVIEWER PRODUCTIVITY

The interviewer productivity ratio divides the number of interviewer minutes spent on the survey by the number of those minutes actually spent in questioning respondents. Large ratios mean proportionately less time spent interviewing respondents and more time spent dialing sample numbers or having preinterview contact with sample households.

Female interviewers have slightly worse overall efficiency ratios than male interviewers (4.0 versus 3.9). The largest effects on productivity of interview data, however, come from other aspects of interviewer duties or other characteristics of the interviewer. We regressed the efficiency ratio on interviewer gender; the proportion of the workload that were refusal conversions (because these are viewed as time-consuming cases), the initial refusal rate (as an indicator of skill in persuasion), and the number of months worked (as net effects of experience on productivity).

$$\text{Efficiency Ratio} = 3.46 + .077 [\text{Gender (1 = Male)}] + .025 [\text{Tenure (Months)}]$$

(.21) (.012)

$$- .070 [\text{Prop. Refusal Conversions}] + 1.09 [\text{Prop. Initial Refusals}]$$

(.46) (.47)

Gender has no statistically significant marginal effect on efficiency ratios, controlling for direct effects of tenure and refusal rates. Interviewers who do refusal conversions are not significantly less productive. The reliable effects once again come from tenure, a variable on which males differ from females, and there is evidence that interviewers with more experience spend proportionately less time in interviews than newer interviewers. This is of direct relevance to issues of relative cost efficiencies of male and female interviewers, as the discussion below notes.

INTERVIEWER COSTS

Interviewer costs are a function of both initial training costs and ongoing salary costs. It is useful to consider the approximately \$300 initial training costs per interviewer as amortized over the complete tenure of the person. Figure 1 documents the higher expected tenure of female interviewers relative to males. More than half of the male interviewers leave within three months; in contrast, six months pass before half of the females have left. Assuming those completing 12 or more months remain on the job an average of 36 months, the expected tenure for males is 7.1 months and for females, 14.3 months. Using an estimate of 20 hours per week employed, this result implies that training costs \$.008 for each minute a male interviewer works, on the average, and \$.004 for females.

We used a similar procedure to estimate ongoing salary costs. Interviewers were hired at an average rate of \$4.05 per hour (\$.0675 per minute); after six months of employment their salaries were raised to \$4.30 (\$.0717 per minute); thereafter, every September (we used every 12 months) they were given an increase that averaged 6% of their current pay rate. This pay schedule rewards experience and produces higher per minute costs for long-term interviewers.

With the above figures we can estimate the total costs per interview minute (see Table 3). For interviewers working more than 13 months the increases in salary costs related to higher tenure exceed the decreases associated with reduced training component costs, so that their total costs per minute on the job are higher. When the efficiency ratios are combined with the total per minute costs, the two gender groups can be compared on the costs per minute spent questioning respondents, that is, the cost of one interview minute. Female interviewers have only slightly higher costs for one minute of interview data than do males (\$.3294 to \$.3097), but such a small difference (about 6%) is likely to occur from sampling error alone.

Using these figures to minimize cost per interview would argue that interviewers with lower tenure (disproportionately male) are

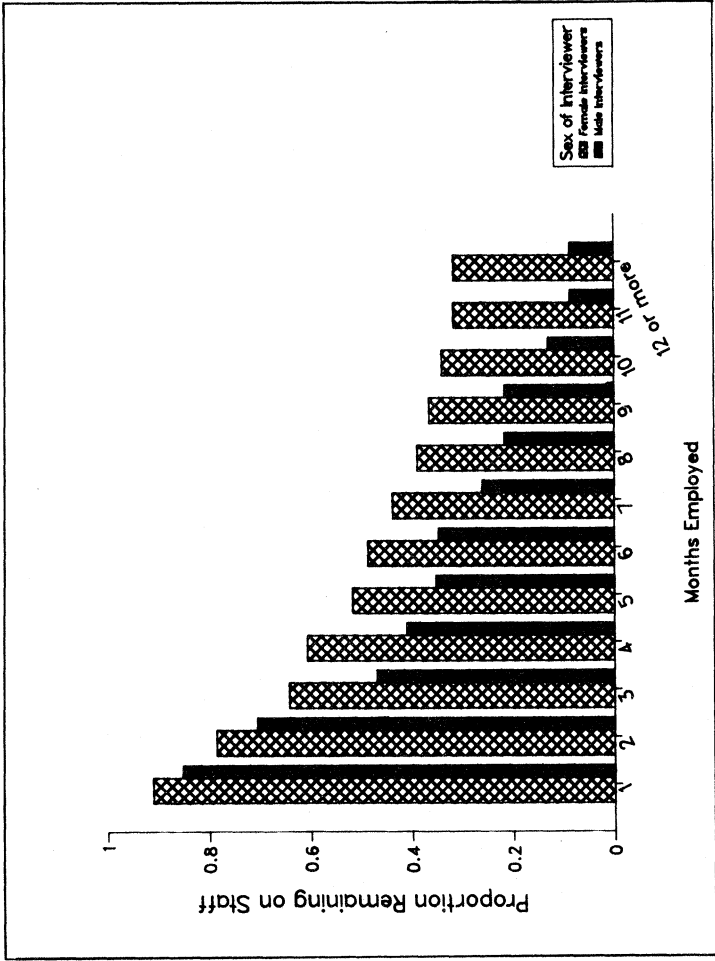


Figure 1: Proportion of Interviewers Remaining on Staff by Month of Employment and Gender

TABLE 3
Estimated Costs per Minute of Interviewer Time
and per Minute of Interview, by Gender

Tenure Group	Per Interviewer Minute Charges			Per Interview Minute Cost
	Training	Salary	Total	
Male Interviewers				
1-6 months	\$.0121	\$.0675	\$.0796	\$.2943
7-12 months	\$.00373	\$.0717	\$.0754	\$.2976
13 or more	\$.00160	\$.0829	\$.0845	\$.3560
Total				\$.3097
Female Interviewers				
1-6 months	\$.00978	\$.0675	\$.0773	\$.2756
7-12 months	\$.00226	\$.0717	\$.0740	\$.2903
13 or more	\$.00160	\$.0829	\$.0845	\$.3644
Total				\$.3294

to be preferred over those with higher tenure. These simple cost comparisons do not reflect, however, the differential skills possessed by interviewers with higher tenure: the higher response rates that they obtain, their ability to shoulder refusal conversion responsibilities that newer interviewers cannot perform well. Cost minimization, in this case, might lead to higher nonsampling errors in the form of nonresponse bias.

DATA QUALITY AND RESPONSE DIFFERENCES

Although males and females appear to be equally cost efficient, there remains the concern that they may collect data of different quality. This could occur in two ways: (1) male and female interviewers may differ in their use of good interviewing techniques,

or (2) independent of the interviewer's behavior, respondents may tend to respond differently to male and female interviewers. We explored both possibilities.

No significant differences were found between male and female interviewers on missing data rates (i.e., the percentages of "don't know" or absent responses). There is also no evidence that male and female interviewers differ in the practice of probing for multiple responses to open-ended questions. For open questions where more than one response is appropriate, the proportions of "second mentions" among male interviewers are nearly identical to those among females.

The Survey of Consumer Attitudes contains a number of non-attitudinal measures, such as the types of vehicles owned by the respondent, the number of cars owned or leased, the model years of these cars, whether they were bought new or used, and the years they were purchased. On the whole, the gender of the interviewer was not significantly related to responses to these questions.

Table 4 shows that respondents express greater optimism toward a variety of economic indicators when the interviewer is male. The differences between the percentages of positive responses recorded by male and female interviewers are significantly large for seven of the nineteen items examined. These seven questions cover diverse economic issues rather than representing a single domain.

ALTERNATIVE EXPLANATIONS FOR INTERVIEWER GENDER EFFECTS

Interviewer gender effects on respondent optimism could greatly complicate substantive inference from these data, and other hypotheses should be explored prior to concluding that they are real. To address these, models were constructed predicting the proportion giving an optimistic response for six variables: whether the respondents felt they would be better or worse off financially in five years, whether they expected their income to be up or down in the next year, whether it was a good or

TABLE 4
 Percentage of Respondents Giving Positive or Negative Response Concerning
 Various Consumer Activities by Gender of Interviewer

Question	Percentage Giving Positive Response			Percentage Giving Negative Response		
	Sex of Interviewer		Standard Error	Sex of Interviewer		Difference
	Male	Female		Male	Female	
Better/worse off now than 1 yr. ago	36.0%	33.7%	2.3%	36.6%	35.9%	0.7%
Better/worse off now than 5 yr. ago	54.3	51.2	3.1	33.9	35.7	-1.8
Better/worse off in 1 yr. than now	36.5	34.6	1.9	14.2	15.5	-1.3
Better/worse off in 5 yr. than now	56.8	51.1	5.7*	13.0	14.1	-1.1
Good/had times financially next 12 mos.	37.5	34.6	2.9	50.7	55.4	-4.7*
Good/had times compared to 1 yr. ago	28.2	26.7	1.5	63.2	64.4	-1.2
Better/worse business cond. in 1 yr.	42.8	38.4	4.4*	17.7	18.9	-1.2
Government economic policy good/poor	21.8	22.2	-0.4	28.2	29.3	-1.1
Less/more unemployed, next 12 mn.	26.3	21.4	4.9*	34.6	38.2	-3.6
Interest Rates in next yr. down/up	39.3	38.2	1.1	25.2	26.2	-1.0
Prices last yr. down/up	6.0	4.5	1.5	72.9	76.8	-3.9*
Prices down/up next 5-10 yrs.	7.2	8.0	-0.8	78.3	74.1	4.2*
Income up more/less than prices last yr.	19.0	16.7	2.3	49.5	50.9	-1.4
Income up more/less than prices next yr.	22.2	18.7	3.5*	34.7	37.0	-2.3
Income up/down next yr.	63.6	59.3	4.3*	15.5	14.7	0.8
Good/had time to buy house	40.5	34.7	5.8*	55.2	60.7	-5.5*
Good/had time to buy durables	55.4	53.1	2.3	34.2	36.7	-2.5
Good/had time to buy car	53.1	48.9	4.2*	39.6	43.1	-3.5
Probably will/won't buy a car	18.3	16.7	1.6	71.6	74.3	-2.7

*Estimate more than twice standard error. Standard errors estimated using a jackknife replication method, treating interviewers as independently selected clusters of observations.

bad time to buy a house, whether business conditions would be better or worse in one year, whether there will be more or less unemployment in the next 12 months, and whether it was a good or bad time to buy a car. The distributions on these variables exhibited large differences between male and female interviewers in Table 4 and a variety of hypotheses were investigated to explain the results.

First, differences in true response values might be confounded with interviewer gender in these data. The proportion of male interviewers was larger during the later months of the study period, when the public was more optimistic about economic conditions. Interviewer gender differences might merely reflect this fact, and to explore this possibility, we controlled for month of survey. The more positive economic attitudes of those interviewed by males remained evident.

Second, there may be interviewer effects on respondents that stem from interviewer characteristics other than gender. We have noted earlier that male and female interviewers differ in the amounts of experience they have with the job. They also might differ in age, education, and other attributes that could have direct effects on respondent reports. For example, the male interviewers tended to be younger than the females. Younger interviewers may administer the questionnaire in a way that elicits more optimistic responses. We investigated whether other interviewer traits account for the response differences by including interviewer gender, age, education, tenure, and response rates in models predicting the consumer attitudes. Despite the fact that sometimes these other variables had independent effects on the consumer attitudes, interviewer gender continued to bear a significant relationship to consumer optimism.

Third, a nonresponse error hypothesis might explain the puzzling finding. Because male and female interviewers tend to gain the cooperation of different kinds of respondents (as evidenced in Table 2), the differences in optimism may really reflect effects of respondent age, gender, and income. Multivariate models including such variables can be used to investigate whether (1) the apparent differences between the results of male

and female interviewers merely reflect the fact that they are interviewing different kinds of respondents, and (2) different magnitudes of gender differences apply among young versus old respondents, male versus female respondents, rich versus poor respondents, and so on. The latter question addresses interaction effects between interviewer gender and various respondent characteristics. Because this hypothesis has strong appeal, we describe a test of it in some detail.

Models were constructed with independent variables reflecting nonresponse differences between the male and female interviewers: respondent gender, respondent age (18-60 versus 61 or over), and respondent income (less than \$10,000 versus \$10,000 or more). Dichotomous independent variables were used in order to provide sufficient stability to the variance and covariance estimates in the multivariate table. Employment status was dropped as an independent variable because it was found to be highly correlated with age and income. Respondent marital status was dropped because of its correlation with age and its low correlations with the dependent variables. Two different models were fit for each set of proportions. The first (the full model) included all direct effects of the independent variables and interaction effects of interviewer gender with each of respondent gender, age, and income. The second model (the trimmed model) dropped all interaction terms.

Table 5 shows that the overall fit of the full model (as judged by probabilities associated with the Wald statistics) is high ($p > .5$) for all variables except the one-year forecasts for business conditions and unemployment. These are the only variables that treat economic conditions external to the respondents themselves, and other independent variables are probably more important for them. Only three of the eighteen different interaction terms reach traditional levels of statistical significance ($p < .05$). The Wald statistics measuring the fit of the model to the data show that only two of the six trimmed models incur important loss of fit due to dropping the interaction terms. This result suggests that differences between male and female interviewers are stable across respondent groups differing on their gender, age, and/or income.

TABLE 5
Parameter Estimates and Their Probability Values for Models Predicting Proportion Positive on Five Consumer Attitudes Using Interviewer Gender, Respondent Gender, Age, and Family Income

Dependent Variable	Selected Full Model Parameters ^a					Trimmed Model ^b	
	Interviewer Gender	Interactions of Interviewer Gender With			Fit of Model ^c	Interviewer Gender	Fit of Model ^c
		Respondent Gender	Respondent Age	Respondent Income			
Better/Worse off in 5 years	.023 (.49)c	-.015 (.64)	.025 (.49)	-.045 (.20)	(.85)	.029 (.08)	(.80)
Income Up/Down next year	.048 (.04)	.0087 (.66)	-.029 (.24)	-.0096 (.68)	(.64)	.029 (.0012)	(.74)
Good/Bad time to buy house	.056 (.126)	-.00019 (1.0)	-.0012 (1.0)	-.016 (.54)	(.73)	.047 (.09)	(.90)
Better/Worse business conditions in 1 year	-.0070 (.83)	.049 (.0051)	.039 (.19)	-.0079 (.75)	(.00)	.051 (.0003)	(.00)
Less/More Unemployment next 12 months	.037 (.11)	.020 (.30)	.0044 (.85)	-.088 (.0002)	(.23)	.0050 (.76)	(.0043)
Good/Bad time to buy car	.049 (.13)	.0080 (.68)	-.034 (.30)	.076 (.02)	(.53)	.054 (.0019)	(.11)

a. Full model specification includes direct effects of interviewer gender, respondent gender, respondent age, respondent income, and interaction effects of interviewer gender and effects of the 3 respondent characteristics.
b. The trimmed model drops all interaction terms.
c. All figures in parentheses are probability values associated with Wald χ^2 statistics, using weighted least squares estimation reflecting clustering of observations into interviewer assignments. For individual coefficients, the probability values are those associated with the null hypothesis of zero coefficient value. For the overall model, the values are those associated with goodness of fit.

Given the lack of influence of the interaction terms, the trimmed model provides the best estimate of the effects of interviewer gender, controlling for respondent gender, age, and income. In five of the six models, the effect of interviewer gender remains strong, with males obtaining significantly more positive answers on the economic attitude questions. (Two of the models have gender effects with probability values on the margin of statistical significance, $p < .10$.) Because these effects are measured in models containing the direct effects of respondent gender, age, and income, they can be interpreted as reflecting the presence of interviewer gender effects within categories of respondents sharing the same sex, age, and income. It is noteworthy that the effects range from 3% to 5% more positive attitudes with male interviewers. These are smaller in general than those without controlling for demographic differences in respondent groups.

In sum, we find little support for the hypotheses that temporal changes in attitudes, other interviewer characteristics, or non-response effects explain the finding of more optimistic answers given to male interviewers. After rejecting alternative explanations, we are left with the possibility that there exist response error differences associated with interviewer gender. That is, respondents answering the questions of a female interviewer behave differently than they would answering the same questions of a male interviewer. The social psychological influences behind these response errors can be hypothesized to lie within interviewer behavior, or the respondent's reaction to the interviewer, or some combination of the two.

On the interviewer side, there is a small literature documenting the effects of interviewer expectations on survey responses. Bradburn and Sudman (1979) and Singer and Kohnke-Aguirre (1979) note that interviewers who, prior to interviewing, believed that respondents would fail to report certain socially undesirable traits did, indeed, obtain fewer reports of such traits. There is, however, little theoretical or empirical description of how interviewer expectations actually affect recorded survey answers. For the consumer sentiment data examined here, one would need to obtain data on interviewers' expectations about survey response

distributions (and on their own economic attitudes) in order to explore this hypothesis.

Viewing the locus of the error as the respondent, several alternative hypotheses could be entertained. If more optimistic economic attitudes are perceived to demonstrate strength, then it is possible that male respondents would be more likely to give these responses to male interviewers in a competitive attempt to appear financially successful. It may be less likely that female respondents would exhibit the same tendency to present a "stronger" facade to male than to female interviewers. Thus this perspective produces an hypothesis of interaction effects of interviewer gender and respondent gender. In Table 5, however, there is no evidence of different magnitudes of interviewer gender effects for male versus female respondents. Hence, this hypothesis is an unlikely explanation of the response errors.

Another source of response error explanatory of the results in Table 4 is that respondents use the gender of the interviewer as an added stimulus to their response formation. Hatchett and Schuman (1975) assert that because white respondents did not want to "offend" black interviewers, they reported racially more liberal attitudes to them than to white interviewers. Respondents, in the absence of strong convictions of their own, may provide answers that they perceive as compatible with those that the interviewers themselves would give. What is to Hatchett and Schuman the avoidance of giving social offense can also be seen as giving answers similar to the interviewer's own attitudes (in their case, attitudes about the interviewer's own racial group).

In this case, it is possible that respondents would ascribe less optimistic economic attitudes to female than to male interviewers. This could arise because of general beliefs about societal differences in economic status between males and females. Alternatively, respondents may assume that a male asking economic questions for the University of Michigan is a professional economist, whereas a female asking similar questions is a clerical person. Either interpretation could lead respondents to believe that the male interviewers would be more optimistic about their own economic future. In the absence of strong attitudes on an

item, the respondents are affected by the ascribed interviewer attitudes.

This explanation is also consistent with the result that fewer gender effects are found on variables concerning changes in economic status over the past year or five years than on attitudes about the future. The response formation process for questions about the past would have the guidance of real experiences of the respondents (e.g., changes in salary levels). The influence of the interviewer on responses may, therefore, be diminished.

Several manipulations of the effect might be attempted as a verification. Status of the interviewers could be communicated directly to the respondent throughout the interview (e.g., at the introduction "I am a research analyst at the university" or "I am a clerical assistant working on this project"). Alternatively, questions for which females are likely to have very different attitudes than males could be used to demonstrate that effects vary by ease of ascription of attitudes toward the interviewers. (This may explain the strong gender effects observed by Ballou and Del Boca, 1980, on surveys of feminist issues.)

SUMMARY, CONCLUSIONS, AND SPECULATIONS

This article has documented the experience of the Survey Research Center over a two-year period covering parts of the years 1981-1983, and used approximately 7,000 interviews taken on the Survey of Consumer Attitudes to address whether male interviewers and female interviewers differ in a large set of characteristics.

We have found that male interviewers tend to terminate their employment sooner than females, and because of this they tend to have lower response rates and higher refusal rates than female interviewers. Males tend to interview proportionally fewer female respondents, fewer older respondents, more people who are working, and fewer poor people. In spite of all these differences, their costs associated with obtaining one minute of interview data are similar to those of females.

Male and female interviewers show no differences in their achieved rates of item missing data or "don't know" answers or in responses to factual questions. On seven of nineteen economic attitude variables males showed a tendency to obtain more positive, or optimistic answers. When several of these measures were investigated in multivariate models, it was found that most of the effects remained in the presence of controls on month of the interview, other interviewer characteristics, and respondent characteristics related to nonresponse differences between male and female interviewers. Differential response error by gender of the interviewer appears to be the most likely cause of the differences.

Although the data restrict our testing of the social psychological underpinnings of the findings, these results have important implications for analysts investigating substantive hypotheses about the structure of economic attitudes. Rarely do analysts of survey data include in their analytic models terms that reflect the type of measurement process, but in this case values of parameter estimates might vary by the gender of the interviewer. This possibility underscores the need for more routine examination of survey data for such response errors and for the mounting of experiments to identify their sources.

REFERENCES

- BALLOU, J. and F. DEL BOCA (1980) "Gender interaction effects on survey measures in telephone interviews." Presented at the 35th annual conference, American Association for Public Opinion Research, Mason, Ohio, May 29-June 1.
- BENNEY, M., D. REISMAN, and S. STAR (1956) "Age and sex in the interview." *Amer. J. of Sociology* 62: 143-152.
- BRADBURN, N. M. and S. SUDMAN (1979) *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- BRILLINGER, D. R. (1964) "The asymptotic behavior of Tukey's general method of setting approximate confidence limits (the jack-knife) when applied to maximum likelihood estimates." *Rev. of the Int. Stat. Institute* 3: 202-206.
- COLOMBOTOS, J., J. ELINSON, and R. LOEWENSTEIN (1968) "Effect of interviewer's sex on interview responses." *Public Health Reports* 83: 685-690.
- DIJKSTRA, W. (1983) "How interviewer variance can bias the results of research on interviewer effects." *Quality and Quantity* 17: 179-187.

- GROVES, R. N. and L. MAGILAVY (1983) "Interviewer effects in telephone surveys." Proceedings of the 1983 Meetings of the International Statistical Institute.
- HATCHETT, S. and H. SCHUMAN (1975) "White respondents and race-of-interviewer effects." *Public Opinion Q.* 39, 4: 523-528.
- HOUSE, J. S. and S. WOLF (1978) "Effects of urban residence on interpersonal trust and helping behavior." *J. of Personality and Social Psychology* 36, 9: 1029-1043.
- INDERFURTH, G. P. (1972) Investigation of Census Bureau Interviewer Characteristics, Performance, and Attitudes: A Summary. Working Paper 34, U.S. Bureau of the Census.
- JOHNSON, W. T. and J. D. DeLAMATER (1976) "Response effects in sex surveys." *Public Opinion Q.* 40, 2: 165-181.
- KENKEL, W. F. (1961) "Sex of observer and spousal roles in decision making." *Marriage and Family Living* (May): 185-186.
- KISH, L. and M. R. FRANKEL (1974) "Inference from complex samples." *J. of the Royal Stat. Society B*, 36, 1: 1-37.
- LANDIS, J. R., J. M. LEPKOWSKI, S. A. EKLUND, and S. A. STEHOUWER (1982) *A Statistical Methodology for Analyzing Data from a Complex Survey: The First National Health and Nutrition Examination Survey*. National Center for Health Statistics, Series 2, No. 92.
- LANDIS, J. R., D. SULLIVAN, and J. SHELEY (1973) "Feminist attitudes as related to sex of interviewer." *Pacific Soc. Rev.* 16: 305-314.
- NEALON, J. (1983) "The effects of male vs. female telephone interviewers," pp. 139-141 in *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- RUMENIK, D. K., D. R. CAPASSO, and C. HENDRICK (1977) "Experimenter sex effects in behavioral research." *Psych. Bull.* 84: 852-877.
- SINGER, E. and L. KOHNKE-AGUIRRE (1979) "Interviewer expectation effects: a replication and extension." *Public Opinion Q.* 43, 2: 245-260.
- STEEH, C. G., R. M. GROVES, R. COMMENT, and E. HANSMIRE (1983) "Report on the survey research center's surveys of consumer attitudes," pp. 173-208 in *Incomplete Data in Sample Surveys*. New York: Academic.
- WAKSBERG, J. (1978) "Sampling methods for random digit dialing." *J. of the Amer. Stat. Assn.* 73: 40-46.

Robert M. Groves is a Senior Study Director at the Survey Research Center and an Associate Professor of Sociology at The University of Michigan.

Nancy H. Fultz is a Research Associate in the Institute of Gerontology at The University of Michigan.