DECISION MAKING OF SURVEY RESPONDENTS

S. James Press and Judith M. Tanur*
Department of Statistics, University of California, Riverside, Riverside, CA 92521-0138 and
Department of Sociology, State University of New York at Stony Brook, Stony Brook, NY 11794-4356

KEY WORDS: Brackets, Confidence, Recall Respondent-Generated Intervals, RGI, Risk, Uncertainty

This research grows out of thinking about a new technique for asking survey questions proposed by Press, 1999. The technique, called Respondent-Generated Intervals (RGI), is designed to be used when the survey task involves recall of a quantity or frequency (What was your income last year? How often have you visited your doctor in the last six months?). The to-be-recalled quantity is referred to as the "usage quantity," and the RGI technique is novel in that it asks the respondent not only to supply a usage quantity, but also to report how low the usage quantity might be (that is, provide a lower bound), and how high the usage quantity might be (an upper bound). From a Bayesian point of view the RGI technique solicits three points on a respondent's subjective recall distribution – some measure of the center of the distribution, and some measures of the extremes. This paper explores how respondents place these three points when asked to report them.

We consider ways in which we might interpret how respondents make their selection of a usage quantity relative to bounds that have already been given, that is, when the bounds question is asked before the usage quantity question. Faced with the decision problem of where in an already-chosen interval they should respond to a recall survey question, how do respondents strategize their responses? The answer relates to how confident the respondent is of his/her response (and how much weight to assign to it).

Some results in economics appear to bear on the issue. Under mild axiomatic conditions of decision-making behavior associated with utility theory, when the rational decision-maker is *completely ignorant*, it has been shown that he/she will opt for one of the extreme choices available; see Arrow and Hurwicz (1972) and Cohen and Jaffray (1980)[1]. This line of research has also been applied in the cognitive psychology context by Hogarth and Kunreuther, (1995); Einhorn and Hogarth (1985, 1986), and Hogarth and Einhorn (1992). Suppose in our context that the respondent has already supplied bounds for the true usage quantity, but he/she is still very uncertain about where in the interval the true usage quantity lies. S/he must now make a decision as to how to report his/her belief about the true usage quantity. If the Arrow/Hurwicz theory is applicable here, it implies that a less confident respondent is likely to opt for a value near one of the boundary points already established. A more confident respondent is likely to report a value for usage closer to the midpoint of the interval. We define a symmetric response to be one at the midpoint of the interval already defined by the lower and upper bounds. In these terms, the least confident respondent is likely to report asymmetrically, and the confident respondent is likely to give a symmetric response.

Some other research in cognitive psychology seems to suggest a diametrically opposite respondent strategy in the case of the bounds questions being asked before the usage quantity question. Referring to a context in which the response intervals are not chosen by the respondent but are pre-set by the interview protocol, Schwarz (1990, p.110) suggests that "respondents may be sensitive to self-presentational concerns when responding. They may be reluctant to check a response alternative that seems extreme in the context of the scale and thus reflects a presumably unusual behavior." See also, Bradburn et al. (1979) and Bradburn and Danis, (1984). Schwarz (1990) also suggests that "…respondents use the range of the response alternatives to infer which behavior is *usual.* In general respondents were found to assume that the behavior of the *average*

*The second author thanks the New York Community Trust for support during the preparation of this paper.

[1] We are grateful to Professor A. Zellner for suggesting these references.

person is represented by the values stated in the middle range of the response scale…." If it is appropriate to adapt the Schwarz theory to the context in which the respondents themselves generate the intervals, we infer that when respondents provide lower and upper bounds, they are creating their own scale, and they then use this scale when they subsequently offer a usage quantity. Accordingly, from this adaptation of the Schwarz theory we conjecture that respondents maximally uncertain about their recall are inclined to give symmetric responses. The Schwarz theory seems to suggest that respondents may attempt to "play it safe" by selecting a usage quantity that is midway between the extremes (the lower and upper bounds). More confident respondents may be inclined to provide asymmetric usage responses that are closer to one of the bounds.

The Arrow/Hurwicz and Schwarz theories appear to have opposite implications for respondent strategy in the factual recall survey question context, with the bounds questions asked before the usage quantity question. Perhaps because the Arrow/Hurwicz axioms are not satisfied in this context, or because the respondent is not completely ignorant (since s/he is assumed to know the true value but merely can't recall it with certainty), the theory really should not be applied. Alternatively, perhaps the Schwarz theory is being over-extended beyond intervals pre-set by the questionnaire protocol, so that in this context the theory is really not applicable. This paper presents empirical data that bear on the question.

The Data

Several record check studies have been mounted to test the accuracy of the RGI technique. We use the data gathered for one of these, more fully described in Marquis and Press (1999) and Press and Marquis (2001) to explore how respondents choose to describe their recall distributions. In this study approximately 500 respondents were interviewed via telephone by Census Bureau interviewers. Questions concerned six items: respondents' salary and wages for the previous year and the year before that, their interest and dividend income for the previous year and the year before that, and their five-year change in salary and wages and in interest and dividend income. There were two versions of the questionnaire; approximately three-quarters of the respondents were asked the bounds questions before the usage quantity question (Version 1), while the other quarter of the respondents were asked the usage quantity question before the bounds questions (Version 2). Respondents to Version 2 of the questionnaire were also asked, for each usage quantity reported, to rate their confidence in their report on a three-point scale (probably very close, probably somewhere between close and far away, probably very far away). For our purposes these responses were coded so that low values of the scale represented high confidence. We will use the results of Version 2 only to establish a method for measuring confidence. Our main analyses will use data from Version 1 where the bounds questions were asked first.

Results

Measuring Confidence

Recall that only in Version 2 (in which the usage quantity was asked first) were respondents asked directly about their confidence in their answers, so our first attempt at looking at the relationship between confidence and symmetry includes only respondents to Version 2. Responses were classified as symmetric if the usage quantity fell exactly at the midpoint of the interval. We found that those who express high confidence in their answers are considerably more likely to offer a symmetric interval than are those who are less confident of their answers. This finding held for 5 of the 6 items, but is hardly to be trusted, for in all cases over half the respondents indicated high confidence, and thus the sample sizes for those who have low confidence are exceedingly low, and those for medium confidence are also often fairly low, making the percents unstable. And no measures of confidence were taken for the majority of respondents, because they were given Version 1 of the questionnaire. We need to seek a proxy for respondent confidence so that we can make the sample sizes closer to equal and so that we can utilize the data for respondents given Version 1 of the questionnaire. Our candidate is relative length of the interval, that is,

Relative length =(upper bound – lower bound)/usage quantity

To see if the relative length variable is a reasonable proxy for the confidence measurement, we ran correlations between relative length and measured confidence for the data from Version 2, shown in Table 1.

These correlations, while not large, are respectable, except for the change variables,

which seem to be very difficult for respondents. There is an issue that the assumptions for use of the correlation coefficient may not be met here, so we also ran Spearman rank order correlations

Table 1 – Correlations between relative length of interval and confidence, non-degenerate intervals only
(Version 2 only)

item	r	n
S/W last yr	.287**	102
S/W prev yr	.247*	106
5-yr Ch S/W	-0.014	95
I/D last yr	.450**	81
I/D prev yr	.319**	83
5-yr Ch I/D	.155	71

**p<.01; *p<.05

and other ordinal measures of association between relative length and confidence. (For gamma, phi, and Somers' d we trichotomized the relative length of the interval.) The results are even more encouraging than the correlational results in Table 1, with the exception of low values for last year's salary and wages. We speculate that these low figures are an artifact of the very skewed distribution of confidence for last year's salary and wages. So we will use relative length of the intervals for our proxy for confidence in what follows.

Placement of usage quantity when the bounds questions are asked first

To see if there was variation in the placement of the usage quantity by relative length of interval (our proxy for respondent confidence), we classified each usage quantity by whether it fell at the midpoint of the interval or not. For Version 1, in which the respondent gave the bounds before giving the usage quantity, Table 2 shows the percent of usage quantities at the midpoint by relative length of interval, trichotomized for this analysis.

Over the 6 items considered in the Press and Marquis study, we see that in all but one case the lowest percentage of usage quantities falling at the midpoint of the interval occurred in the longest relative interval, where presumably the respondents had the least confidence in their answers. We calculate the probability of these 5 successes (or more) out of 6 trials, with probability of success on each trial equal to 1/3, to be less than .002.[2]

Table 2: Percents of usage quantities falling at midpoint by relative length of interval (Version 1 data only – numbers in parentheses are base n's for percentages)

	Interval Length		
	Short	**Med**	**Long**
S/W last yr	38% (104)	27% (100)	10% (104)
S/W prev. yr	35% (108)	31% (108)	16% (109)
S/W 5-yr Ch.	46% (99)	36% (88)	21% (94)
I/D last yr	51% (73)	44% (78)	20% (83)
I/D prev. yr	47% (78)	38% (79)	22% (80)
I/D 5-yr Ch.	56% (81)	29% (65)	30% (73)

[2] This probability calculation is at most suggestive, as it assumes that the trials are independent. But we have reason to believe that they are not independent, as people seem to form habits about the relative placement of their usage quantity within their intervals. Table 3 illustrates this point for Version 1 of the questionnaire.

Table 3: Correlations between relative positions of usage quantities (u-lb)/(ub-lb)

	S/W Prev yr	**S/W 5yr Ch**	**I/D Last yr**	**I/D Prev yr**	**I/D 5yr Ch**
S/W last yr	.471**	.172	.150	.134	.098
S/W prev yr		.021	.233**	.146*	-.031
S/W 5-yr Ch			.066	.063	.097
I/D last yr				.478**	.175*
I/D prev yr					.294*

*p< .05
**p<.01

What we see in this table is the correlation across items of the relative position chosen for the usage quantity in the interval, computed as (usage – lower bound)/(upper bound – lower bound). We see that some of these correlations, especially those between the same kinds of income, are considerable and hence the independence assumption is untenable.

Further, we see that in 5 of the 6 items there is a pattern of the percent of usage quantities falling at the midpoint decreasing monotonically with increasing relative length of the intervals. If we take the probability of that permutation occurring by chance to be 1/6, we calculate that the chances of 5 or more such successes out of 6 items to be less than .0002.[3] These findings favor the Arrow-Hurwicz hypothesis rather than that of Schwarz.

It seemed to us that the requirement that a usage quantity fall exactly at the midpoint of the interval was a very stringent definition of a "symmetric" interval, so we calculated Table 3 to relax that definition. Although it would be realistic to assume that respondents' recall distributions for items of income would be Pareto in form, for simplicity we returned to some earlier thinking that treated such distributions as approximately normal. This is merely a matter of convenience in order to construct the zones explained below. According to that formulation, if respondents really constructed their intervals to cover virtually all possible values of the usage quantity, and if we assume the respondents' subjective usage distribution was approximately normal, then the length of an interval would represent 6 standard deviations of that respondent's recall distribution. Hence we divided the length of the interval by 6 to estimate the standard deviation of the respondent's recall distribution, sigma. We then laid out 5 zones – (i) more than 2 sigma below the midpoint of the interval, (ii) between .67 and 2 sigma below the midpoint, (iii) between .67 sigma below the midpoint and .67 sigma above the midpoint, (iv) between .67 and 2 sigma above the midpoint, and (v) above 2 sigma above the midpoint. Note that if the normality assumption holds, then zones (i) and (v) represent the lower and upper 2.5% tails of the respondent's recall distribution, respectively, and zone (iii) represents the middle 50% of the respondent's recall distribution. We then tabulated the percent of usage quantities falling into the middle 50% of the distribution by relative length of interval. These results appear in Table 4.

Table 4 shows the same pattern as Table 2, with all 6 of the items showing the lowest percent symmetric intervals in the case of the longest intervals, and all of the patterns being monotone decreasing.

Table 4: Percents of usage quantities falling in middle 50% zone of distribution by relative length of interval (Version 1 data only – numbers in parentheses are base n's for percentages)

	Interval Length		
	Short	**Med**	**Long**
S/W last yr	51% (104)	43% (100)	35% (104)
S/W prev. yr	53% (108)	45% (108)	33% (109)
S/W 5-yr Ch.	56% (99)	54% (88)	35% (94)
I/D last yr	67% (73)	58% (78)	42% (83)
I/D prev. yr	60% (78)	54% (79)	41% (80)
I/D 5-yr Ch.	64% (81)	42% (65)	41% (73)

Conclusion

When the bounds questions were asked first, we found that respondents presumed to be least confident in their recall, that is, those who had long relative length of intervals, were more likely than those more confident to place their usage quantities towards the extremes of the intervals. This finding held whether we defined the "extremes" as anything other than the midpoint of the interval or as what could be the upper and lower 2.5% tails of the respondent's recall distribution if that distribution were normal. This evidence suggests that the self-generated bounds intrinsic to the RGI protocol afford respondents freedom to locate their responses according to their own degree of confidence. This may reduce bias compared to the effects of the normative constraining influence that Schwarz has found for interviewer- or questionnaire-supplied intervals. We further conclude that the evidence presented here supports the Arrow/ Hurwicz position that rational decision makers in a state of ignorance will make extreme choices.

References

Arrow, K. J. and L. Hurwicz (1972). "An Optimality Criterion for Decision Making Under

[3] Again, this probability calculation is merely suggestive.

Ignorance" in C. F. Carter and J. L. Ford (Eds.) *Uncertainty and Expectations in Economics*, Oxford: Basil Blackwell, pp. 1-11.

Bradburn, Norman; Sudman, S. and Associates (1979). *Improving Interviewing Methods and Questionnaire Design*, San Francisco: Jossey-Bass.

Bradburn, Norman and Danis, C. (1984). "Potential Contributions of Cognitive Research to Survey Questionnaire Design," in *Cognitive Aspects of of Survey Methodology: Building a Bridge Between Disciplines*, T.B. Jabine, M.L. Straf, J.M. Tanur, and R. Torangeau (Eds.), Washington, DC: National Academy Press, pp.101-129.

Cohen, M. and Jaffray, J. (1980). "Rational Behavior under Complete Ignorance," *Econometrica*, **48**, 1281-1299.

Einhorn, H. J. and Hogarth, R. M. (1985). "Ambiguity and Uncertainty in Probabilistic Inference," *Psychological Review*, **92**, 433-461.

---------- (1986). "Decision Making under Ambiguity," *Journal of Business*, **59** (4), Part 2, S225-S250.

Hogarth, R. M. and H. J. Einhorn (1992). "Order Effects in Belief Updating: The Belief Adjustment Model," *Cognitive Psychology*, **24**, 1-55.

----------and H. Kunreuther (1995). "Decision Making under Ignorance: Arguing with Yourself," *Jour. of Risk and Uncertainty*, **10**, 15-36.

Marquis, Kent H., and Press, S. James (1999). "Cognitive Design and Bayesian Modeling of a Census Survey of Income Recall," *Proceedings of the Federal Committee on Statistical Methodology Conference*, Washington, DC, Nov. 16, 1999, pp.51-64 (see http://bts.gov/fcsm).

Press, S. James (1999). "Respondent-generated Intervals for Recall in Sample Surveys," manuscript, Department of Statistics, University of California, Riverside, CA 92521-0138 , Jan., 1999. http://cnas.ucr.edu/~stat/press.htm

Press, S. James and Marquis, Kent H. (2001) "Bayesian Estimation in a U.S. Government Survey of Income Using Respondent-Generated Intervals." *Research in Official Statistics*. **4**, 151-168.

Schwarz, Norbert (1990). "Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction", *in Research Methods in Personality and Social Psychology*, C. Hendrick and M.S. Clark (Eds.), (*Review of Personality and Social Psychology*, Vol. 11, pp.98-119), Beverly Hills, CA: Sage Publications.

Schwarz, Norbert, Sudman, Seymour (Eds.) (1994). *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer-Verlag.