Little Things Matter:

A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses

Tom W. Smith

National Opinion Research Center
University of Chicago

GSS Methodological Report No. 78

July, 1993

This research was done for the General Social Survey project directed by James A. Davis and Tom W. Smith. The project is supported by the National Science Foundation, Grant No. SES-9122462.
I would like to thank Woody Carter and David Mingay for comments on drafts of this paper.

It is well known that seemingly minor changes in question wording, response format, and context can appreciably alter response distributions. What is less appreciated is that non-verbal aspects of surveys such as physical layout and visual presentations can also notably influence answers. Below we cite five examples where variations in such matters affected how interviewers, respondents, or both handled and responded to questions:

1) Misalignment of Response Categories
2) Dutch Ladders
3) Placement of Follow-up Questions
4) Overly Compact Question Formats
5) Open-ended Questions and Wide Open Spaces

Misalignment of Response Categories

The 1993 International Social Survey Program (ISSP) study on the environment was administered as a self-completion supplement to NORC's General Social Survey (GSS) (Davis and Smith, 1992). Due to a font problem the final master of the questionnaire misaligned the response boxes to Q.21b. The boxes were pushed one tab to the right so that the left-hand box appeared where the right-hand box should have been and the right-hand box was shifted into the right margin (See Figure 1). This error was discovered when the questionnaires were returned from the printer. No correction was made since it was assumed that the intent of the response categories was clear and that respondents would mark the appropriate box even though misaligned. This was not the case.

The misaligned boxes confused many respondents. First, the number of No Answers increased dramatically. For six items immediately before or after Q.21 No Answers ranged from 30-35 (average 32.8). For Q.21a No Answers more than doubled to 70 and for the misaligned Q.21b more than quadrupled to 134. Q.21c had 121 No Answers. Moreover, this increase was related to educational level and verbal ability. On five preceding items on environmental actions there was no association between education or verbal ability (measured by a 10-item vocabulary test) and giving No Answers. However, for the misaligned Q.21b and the following item giving No Answers was significantly related to less schooling and lower verbal ability. For example, on Q.21b 14.1% of those with less than a high school education had No Answer as opposed to 8.4% of high school graduates, 7.6% of college graduates, and 6.1% of those with advanced degrees. Those with less education and verbal skills were most affected by the confusing layout.

Second, many respondents who checked box 1 meant "No" rather than "Yes". They apparently followed the vertical alignment of boxes reasoning that box 1 meant "No" since it was physically underneath the "No" header. (Rather than meaning "Yes" as the first or left-hand response option.) We believe this to be the case mainly because other recent surveys on contributing to environmental groups produce consistently lower estimates of giving

FIGURE 1

OFFICE USE ONLY

19d. And how often do you **cut back** on driving a car for environmental reasons?

*PLEASE CHECK **ONE** BOX ONLY*

	(✓)
Always	1
Often	2
Sometimes	3
Never	4
I do not have or cannot drive a car	5

---

20. Are you a member of any group whose main aim is to preserve or protect the environment?

*PLEASE CHECK **ONE** BOX ONLY*

	(✓)
Yes	1
No	2

---

21. In the last **five years,** have you...

*PLEASE CHECK **ONE** BOX ON EACH LINE*

	Yes, I have	No, I have not
a. ...signed a petition about an environmental issue?	1	2
b. ...given money to an environmental group?	1	2
c. ...taken part in a protest or demonstration about an environmental issue?	1	2

Please continue ...

than the 1993 GSS. With No Answers excluded 70.9% of GSS respondents indicated they had given money within the last five years (i.e. were coded as in box 1). Five similar (but not identical) questions asked by Gallup, Wirthlin, Gordon Black, and Opinion Research Corporation from 1988 to 1992 showed giving rates of 36-49% while two other 1990 surveys by Yankelovich and Hart-Tetter indicated that from 38 to 51% never give money to environmental groups. Taken together these alternative estimates suggest that the GSS numbers are too high by 15-20 percentage points. In addition, we looked at how membership in an environmental group (Figure 1 - Q.20) related to the giving question. We physically examined about 10% of the questionnaires of those who belonged to an environmental group and who gave money. Of these 46% had either drawn in a box in its proper location or placed a check in this same location. Of those who did not belong to an environmental group, but who gave money, only 15% drew a box or placed a check where box 1 should have been. This pattern suggests that giving by non-members of environmental groups may have been exaggerated (assuming that being a member is unrelated to clarifying one's response to Q.21b by placing one's answer in the physically correct position.)

As a result, we can be reasonably certain that "No" responses mean exactly that, but only for the approximately 20% of "Yes" responders who physically clarified their response can we be sure they meant "Yes". The remaining 80% who checked box 1 without elaboration consist of a mixture of givers and non-givers.

Dutch Ladders

The 1987 ISSP study on social inequality included a measure of subjective social stratification:

> In our society there are groups which tend to be towards the top and groups which tend to be towards the bottom. Below is a scale that runs from top to bottom. Where would you place yourself on this scale.

There were 10 response categories with 1=Top and 10=Bottom. This item was asked in nine countries (Australia, Austria, Germany (West), Great Britain, Hungary, Italy, the Netherlands, Switzerland, and the United States). All countries show a majority placing themselves towards the middle (4-7), but the Netherlands clearly is an outlier (Table 1). The range in the % placing themselves in the middle is 24.0 percentage points from 83.8% in Australia to 59.8% in the Netherlands. Over half the overall difference (12.4 percentage points) is due to the Netherlands. Likewise, at the bottom (8-10) the range is 31.3 percentage points with the Netherlands contributing almost half (13.6 percentage points). While most of the other differences appear to reflect actual differences in social structure, the Netherlands' distinctive distribution does not fit other measures of Dutch society (e.g. income distributions), nor is the Netherlands so

distinctive on other social inequality measures (e.g. subjective class identification) (Smith, 1990).

This raised translation as a likely suspect for the Dutch deviation, but an examination of the Dutch wording indicated it was equivalent to the English in meaning and appropriate and clear in Dutch. It was then discovered that the visually displayed scale in the Netherlands differed from that employed in the other countries. The intended scale was to have 10 vertically stacked squares (with the highest box labelled "Top" and the lowest labelled "Bottom"). The Dutch scale had 10 stacked boxes, but they were in the shape of a truncated pyramid, with the bottom boxes wider than those in the middle and top. It appears that Dutch respondents were attracted to the lower boxes because they were wider and were probably seen as indicating where more people were.

Table 1

Distribution of the 10-Point Social Rank Question

	"Top" 1-3	"Middle" 4-7	"Bottom" 8-10
Australia	10.4%	83.8	5.8
Italy	9.9%	83.6	6.6
Germany	9.8%	80.9	9.2
United States	17.6%	72.2	10.1
Switzerland	11.2%	77.9	10.9
Austria	6.0%	79.5	14.3
Great Britain	7.7%	75.2	17.1
Hungary	2.5%	74.0	23.5
The Netherlands	3.2%	59.8	37.1

Source: 1987 ISSP

Placement of Follow-Up Questions

Skips (i.e. questions/instructions that tell interviewers to ask different follow-up questions depending on responses to prior questions) are relatively hard for interviewers to correctly follow. In paper and pencil questionnaires various devices such as skip instructions, arrows, flags (e.g. circled question numbers, pointing fingers), and IQ boxes (boxes with questions that interviewers must answer before preceding) are used to guide interviewers. An example of a skip or filtered question is the GSS items on religion and religious strength.

The GSS question on religion consists of two parts. Everyone is asked their main religion (Protestant, Catholic, Jewish, Other, or None). Protestants and Jews are then asked their denomination or branch. After this follow-up Protestants, Catholics, Jews, and Others are asked "Would you call yourself a strong [PREVIOUSLY NAMED RELIGION FROM PRECEDING QUESTION] or a not very strong

[PREVIOUSLY NAMED RELIGION FROM PRECEDING QUESTION]? Those with No Religion skip over this follow-up question.

The proportion giving No Answer to the religious strength follow-up question (almost all due to failure to ask by interviewers) has varied considerably in recent years from a low of 1.7% in 1987 to a high of 11.8% in 1988 (1985-91 average=5.2%). Since wording, instructions, skip patterns, and order are unchanged across these years most sources of variation do not come into play.[1] However, the physical placement of the follow-up item did vary. It variously appeared at the bottom of the same page as the religion question, at the top or middle of the following right-handed page (i.e. facing the religion question), and at the top or middle of the following left-hand page (i.e. on the backside of the page with the religion question).

It appears that the physical placement of the religious strength item was one factor contributing to the variation in No Answer levels. When it appeared at the top of a backside page, this increased incorrect interviewer skips. This is shown most clearly in 1988. Three ballots (i.e. versions of the questionnaire each given to a random third of the sample) were used on the 1988 GSS. On two ballots the religious strength item appeared at the top of a backside page. On the third the item appeared at the top of a facing page. The % missing for the two backside ballots was 14.4% and 15.2%, while for the facing ballot it was 6.1%. The difference was statistically significant at the .0000 level. Differences across other years and ballots show the same pattern, but not so strikingly.

Overly Compact Question Formats

On the 1972 and 1973 GSSs the four educational attainment questions (self, spouse, mother, and father) were placed in a grid format on one page. The educational sub-questions ran down the side and the four persons were listed across the top. This dense format produced a high number of No Answers (presumably errors in interviewers following skips). For example, in 1972 and 1973 the % No Answer for respondent was respectively 1.4% and 1.0% and for father was 9.0% and 6.9%. In 1974 and subsequent years the questions were each placed on a separate page. In 1974 and 1975 the % with No Answer was less than 0.1% for self and 0.4-1.2% for father.

Open-ended Questions and Wide Open Spaces

Allowing more space for recording open-ended answers apparently produces longer recorded responses and perhaps responses

---

[1]With the exception that the follow-up question for Jews was asked in 1988-1991, but not in 1985-87. In 1985-87 only Protestants were asked their denomination while Catholics, Jews, and Others went directly to the religious strength item.

closer to actual verbatims. The 1954 Stouffer study on civil liberties and Communism was jointly fielded by Gallup and NORC using a common questionnaire. Each organization separately printed its own copies and on open-ended questions NORC allowed five times as much open space for recording answers as Gallup did (Stember, 1955). A word count of responses to two questions showed means of 13.6 and 13.7 words for Gallup and 23.6 and 18.4 words for NORC (Stember, 1955). While different interviewing staffs may also explain these differences, it is likely that allotting more space for answers both facilitates and encourages the recording of longer and more detailed answers.

Summary

Both respondents and interviewers can be affected by physical layout and other visual aspects. Respondents were seriously confused by a seemingly simple misalignment of response categories relating to giving money for the environment and Dutch response to a social status item was notably shifted because of a different presentation of a 10-point scale. Similarly, interviewers' performance and accuracy can be appreciably affected by the physical layout. On items dealing with religion and education the frequencies of No Answer responses resulting from difficulties in following skip patterns differed because of the layout of the questions. In addition, the amount of open-ended material that interviewers record apparently depends in part on the amount of physical space allotted.

These findings parallel those from the educational testing field which shows that differences in the answering mode (e.g. circling a letter besides multiple choice responses vs. filling-in ovals on an answer sheet) and test booklet format can notable affect test scores (Beaton, 1988; Hedges, 1989; Earles, Guiliano, Ree, and Valentine, 1983; Hilton, 1992; Rock, et al., 1985).[2]

Similarly, in psychological experiments of context and conversational norm (Strack, Schwarz, and Waenke, 1991; Schwarz, Strack, and Mai, 1991) layout has been used to join together or separate adjoining items (e.g. by surrounding two questions in a box vs. separating them on different pages or even in different questionnaires). These differences in layout apparently lead to different connections and comparisons being made between items and to significant changes in correlations.[3]

The above survey examples and the educational testing and

---

[2]I would like to thank Steven Ingles of NORC for introducing this literature to me.

[3]Since these experiments also varied labels as well as layout one cannot be sure that layout is contributing to the observed effects. However, layout was manipulated in each experiment to produce just the result that were observed and it probably was a significant factor.

psychological research indicate that close attention must be given to physical layout and other "trivial" visual matters when questionnaires are designed. Without such attention data quality can seriously suffer and replication can be undermined. In surveys as in life little things matter.

References

Beaton, Albert E., The NAEP 1985-86 Reading Anomaly: A Technical Report. Princeton, NJ: Educational Testing Service, 1988.

Davis, James A. and Smith, Tom W., The NORC General Social Survey: A User's Guide. Newbury Park, CA: Sage Publications, 1992.

Earles, J. A.; Giuliano, T.; Ree, M.J.; and Valentine, L.D., The 1980 Youth Population: An Investigation of Speeded Subtests. Brooks Air Force Base, TX: Air Force Human Resources Laboratory, 1983.

Hedges, Larry V., "The NAEP/ETS Report on the 1989 Reading Data Anomaly: A Technical Critique," in Report of the NAEP Technical Review Panel on the 1986 Reading Anomaly, the Accuracy of NAEP Trends, and Issues Raised by State-Level NAEP Comparisons. National Center for Education Statistics Technical Report. Washington, D.C.: NCES, 1989.

Hilton, Thomas L., "Pooling Results from Two Cohorts Taking Similar Tests, Part 1: Dimensions of Similarity," in Using National Data Bases in Educational Research, edited by Thomas L. Hilton. Hillsdale, NJ: Lawrence Erlbaum, 1992.

Rock, D.A.; Hilton, T.L., Polack, J., Ekstrom, R.B., and Goertz, M.E., Psychometric Analysis of the NLS and the High School and Beyond Test Batteries. Washington, D.C.: National Center for EDucation Statistics, 1985.

Schwarz, Norbert; Strack, Fritz; and Mai, Hans-Peter, "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis," Public Opinion Quarterly, 55 (Spring, 1991), 3-23.

Smith, Tom W., "Social Inequality in Cross-National Perspective," in Attitudes to Inequality and the Role of Government, edited by J.W. Becker, James A. Davis, Peter Ester, and Peter P. Mohler. Rijswijk: Sociaal en Cultureel Planbureau, 1990.

Stember, Charles Herbert, "The Effect of Field Procedures on Public Opinion Data," Unpublished Ph.D. dissertation, Columbia University, 1956.

Strack, Fritz; Schwarz, Norbert; and Waenke, Michaela, "Semantic and Pragmatic Aspects of Context Effects in Social and Psychological Research," Social Cognition, 9 (Spring, 1991), 111-125.