An investigation of the frequency and correlates of primacy effects in questions using show-cards in face-to-face interviews

Patten Smith, Louise Sullivan, Hayk Gyuzalyan and India Tracy
Ipsos MORI

Introduction

A commonly used question form in social surveys requires respondents to choose one or more answers from a list of categorical response options presented either visually (on a show-card or self-completion questionnaire) or orally by an interviewer. It has been known for many years that such questions are susceptible to response order effects (Rugg and Cantril, 1944; Payne 1951; Schuman and Presser, 1981; Krosnick and Alwin, 1987). Two types of response order effects are generally recognised: primacy effects, where respondents are more inclined to pick items near the top of the list and recency effects where respondents are more inclined to pick items near the bottom of the list. The more recent literature on the subject has focused on developing theoretical explanations for observed effects (Krosnick, 1991; Krosnick and Alwin, 1987; Narayan and Krosnick, 1996; Schwartz et al, 1991; Tourangeau et al, 2000; Duffy, 2003; Knauper 1999a; Knauper 1999b), rather then on documenting the overall scale of these effects. It is, however, apparent from the literature that such response order effects often *do* occur, that sometimes they are substantial, but also that often they do *not* occur (Payne 1951; Schuman and Presser, 1981; Narayan and Krosnick, 1996). This presents the survey practitioner with something of a quandary: how worried should (s)he be that response order effects will occur when writing questions?

Response order effects occur when respondents are asked to pick answers from lists. Such questions come in several forms. The literature distinguishes between cases where respondents are presented with response options visually (eg on a show-card) or orally by an interviewer. Visually presented items are generally found to be susceptible to primacy effects whereas orally presented items are more susceptible to recency effects (Krosnick, 1991; Krosnick and Alwin, 1987). A second important distinction, which cross-cuts mode of presentation relates to *response coding instructions*, by which we mean the number of options the respondent is asked to select. Respondents may be asked to select as many items as they feel are appropriate (the *code all that apply* format), they may be asked to pick one only (the *code one only format*) or they may be asked to select a set number, or maximum number, of items where this number is greater than one (the *code n items* format). Although this second distinction is well recognised by the survey practitioner and presents rather different task demands to respondents, it has not been developed in the literature.

The purpose of this paper is three-fold. First, we investigate the extent to which primacy effects were observed in questions which visually presented response options in a 12 month series of Ipsos MORI omnibus surveys. Second, we investigate the extent to which the frequency of primacy effects varies by response coding instructions. Third, we use the limited data we have available to test a number of predictions about response order effects which can be taken from the theory most commonly used to try to explain primacy effects, namely the theory of satisficing.

Satisficing theory (Krosnick, 1991) attempts to account for response effects by arguing that under certain circumstances, in answering a question respondents may choose to provide a merely *satisfactory* answer rather than an *optimal* one in order to avoid the substantial cognitive effort entailed in producing the latter. In an ideal world, respondents would 'carefully interpret the meaning of each question, search their memories extensively for all relevant information, integrate that information carefully into summary judgements, and report those summary judgements as clearly and precisely as possible' (Krosnick, 1991). Satisficing theory argues, however, that in reality respondents do not always go through

those four processes thoroughly, as these require substantial effort for little perceived reward. They take short-cuts. The process of taking such short-cuts with the aim of reducing cognitive effort is termed satisficing.

Krosnick argues that satisficing accounts for a wide range of recognised response order effects including primacy effects with visually presented response options. A primacy effect is defined as occurring when respondents preferentially select items placed at the beginning of a list presented to them. Satisficing theory argues that survey respondents, who are motivated to save time and cognitive effort when answering questions, will not carefully consider each item on an unordered list, but rather will read from the top of the list, and will choose the first satisfactory response or responses they encounter (Krosnick and Alwin, 1987; Tourangeau et al 2000).

Additionally, Krosnick and Alwin (1987) suggested that, in addition to satisficing, two other processes may be implicated in producing primacy effects with visually presented response options. First, respondents are likely to process early items more deeply than later ones: their minds will be "cluttered with thoughts about previous alternatives that inhibit consideration of later ones" (Krosnick and Alwin, 1987, p.203), and, because respondents tend to use confirmatory strategies in evaluating options, this greater attention to earlier items will tend to increase their likelihood of being selected. Second, options presented at the beginning of the list may establish a standard of comparison against which subsequent options are judged.

Krosnick (1991) has proposed that three conditions will affect the extent to which satisficing strategies will be implemented by respondents:

- the difficulty of the task: itself a function of the way the question is worded, its clarity, the difficulty of the retrieval process involved (eg the time-point at stake, or the numbers of objects respondents are asked about), and the difficulty of the judgment task required (e.g. ranking objects is found to be more difficult than simple rating scales);
- the respondent's ability to perform complex cognitive tasks;
- The respondent's motivation to optimise.

In the work reported in this paper we identified all questions in a series of Ipsos MORI omnibus surveys that asked respondents to pick one or more answers from unordered lists presented on show-cards, and identified whether or not a primacy effect was evident in each. Analyses were conducted separately for questions with different response coding instructions. We then conducted regression analyses designed to ascertain whether variables which might be expected to be related to task difficulty and respondent motivation were related to the occurrence and extent of primacy effects in the manner predicted by satisficing theory.

The study

In summary the work involved the following steps:

1. identifying all questions in 12 implementations of the Ipsos MORI omnibus in which (i) respondents were asked to select answers from an unordered list on a show-card (ii) the list was presented in reverse order to a random half of the respondents and (iii) at least 100 respondents were asked the question in each implementation order;
2. constructing a dataset of questions containing, for each question, the proportion of respondents who picked an item when it was presented first and when it was presented last;
3. undertaking a simple meta-analysis to assess (i) the number of questions exhibiting primacy effects in answers to first/last items (ii) the magnitude of such effects and

(iii) which question features were associated with propensity to generate primacy effects.

The CAPI Omnibus survey is a face to face survey of approximately 2000 respondents carried out regularly by Ipsos MORI. There are a number of core questions asked at each wave, such as party political support and technologies used. The remaining questions are paid for by different organisations and tend either to appear one wave only or, in the case of tracker items, appear at set time intervals. The 12 implementations of the Omnibus survey covered the period June 2006 to December 2006.

147 question implementations were identified in which an unordered list of response codes was presented on a show-card and in which the list was presented in reverse order to a random half of the respondents. Forty of the 147 question implementations used questions that were fielded in more than one round of the omnibus, and 107 were fielded once only.

For each question implementation four percentages were recorded:

1. percentage picking first response code when card was presented in normal order implementation;
2. percentage picking same response code when presented last[1] in reverse order implementation;
3. percentage picking last response code when card was presented in normal order implementation;
4. percentage picking same response code when presented first in reverse order implementation.

For each question we then calculated four indicators of primacy effects:

1. indicator of presence of primacy effect for first item in normal presentation order: a primacy effect was coded as occurring if the percentage picking this code was higher in normal order than in reverse order;
2. indicator of presence of primacy effect for first item in reverse presentation order: a primacy effect was coded as occurring if the percentage picking this code was higher in reverse order than in normal order;
3. primacy effect *effect size* for first item in normal presentation order: percentage picking this code in normal order minus percentage picking it in reverse order;
4. primacy effect *effect size* for first item in reverse presentation order: percentage picking this code in reverse order minus percentage picking it in normal order.

Questions were classified on the basis of the following:

1. response coding instruction (ie the number of codes respondents were instructed to pick): *code one only*, *code all that apply* or *code n items*[2];
2. number of words in the question stem;
3. number of substantive response codes appearing on the showcard (excluding 'other', 'don't know', 'refused', etc);
4. position of question in questionnaire: first half or second half;
5. percentage of respondents picking the item (mean of percentages when presented first and last).

As discussed in the introduction, we felt that it was important to recognise the distinction between response coding instructions. The response task differs somewhat by coding

[1] When showcards included response codes for ‘other’ answers, ‘don’t know‘ or ‘refused answer’ these always appeared at the bottom of showcards in both normal and reverse order implementations; because these response codes were of no substantive interest for this work, and in any case were not subject to the two order implementations, they were excluded from analysis.

[2] See Appendix for examples of each question type.

instruction and we therefore felt it was worth exploring whether coding instruction affected a question's susceptibility to primacy effects. We also note that task difficulty is likely to vary by response coding instruction. *Code all that apply* questions would appear to be the simplest as all the respondent has to do is sequentially check applicability of each item. The task is harder for *code one only* questions because there the respondent has, not merely to check applicability of each item, but also to compare it against all the other items in order to decide which item is the *most* applicable. Answering *code n items* questions are, arguably, still harder to answer because they involve respondents in making the same kinds of item comparisons as they do for *code one only* questions, but this time repeatedly (first to identify the most applicable item, then the second most applicable item, etc). For this reason we would predict from the theory that the greatest amount of satisficing would occur with *code n items questions* and the least with *code all that apply* questions. In other words questions of the former type would be expected to most susceptible to primacy effects and questions of the latter type would be expected to be least susceptible.

We would also expect number of words in the question stem and the number of response options to be related to task difficulty. According to the theory we would therefore predict that questions with longer stems and more response codes were more susceptible to primacy effects.

Position in the questionnaire might be expected to be related to respondent motivation. Respondents answering questions later in the questionnaire may be more fatigued and less enthusiastic about using optimising strategies when answering questions that respondents answering questions earlier in the questionnaire. We would therefore predict that questions in the second half of the questionnaire would be more susceptible to primacy effects than questions in the first half.

The percentage of respondents picking an item can be seen as an index of its plausibility. According to satisficing theory, satisficing respondents will pick *plausible* items until they deem that a satisfactory answer has been given. Given this, less plausible items would be expected to be less susceptible to satisficing and primacy effects.

Results

All analyses reported here have been conducted separately for (i) items that are first in the list in normal presentation order and (ii) items that are first in the list in reverse presentation order because analysing them jointly would have violated the assumption that observations should be independent [3].

The frequency of primacy effects

Table 1 shows the proportion of questions for which primacy effects were observed broken down by response coding instructions.

Because we have reason from both the empirical and theoretical literature to predict the presence of primacy effects, we have applied one-tailed statistical tests in this table and in table 2.

[3] Strictly speaking we also violate this assumption by including more than one question from each of the 12 omnibus surveys. Given the large omnibus samples and the disparate nature of the questions included in the analysis, we consider it very unlikely that this will have had any substantial effect on the analyses presented here.

Table 1 Frequency of primacy effects by response coding instructions

	Primacy effect	No primacy effect	Base (no. questions)	P under null hypothesis (1-tailed)
Normal order				
All questions	85	61	146*	P<0.05
Code one	28	10	38	<0.01
Code all that apply	30	48	78	NS
Code n items	27	3	30*	<0.01
Reverse order				
All questions	103	44	147	<0.01
Code one	24	14	38	NS (<0.1)
Code all that apply	52	26	78	<0.01
Code n items	27	4	31	<0.01

* one question was excluded from analysis where same percentage picked code same in the two implementation orders

One-tailed binomial tests on findings for the complete set of questions show that respondents were significantly more likely to pick items when presented first than when presented last for items presented first in both normal and reverse order. Thus we have clear evidence that primacy effects occur.

When we looked at the findings for questions with different response instructions findings were mixed. *Code n items* questions exhibited clear primacy effects both for items presented first in normal order and for items presented first in reverse order. Significant primacy effects were found for questions in which respondents chose one answer only (*code one only* questions) for items presented first in normal order and marginally significant primacy effects ($p<0.1$) were found for items presented first in reverse order.

Significant primacy effects were found for items in which respondents picked as many answers as they wished (*code all that apply* questions) for items presented first in reverse order, but not for items presented first in normal order. Indeed, for the latter items significantly more respondents picked codes when they were presented *last* than when they were presented first ($P<0.01$, 2-tailed); in other words for these items we have evidence of the presence of *recency* effects.

In summary we have clear evidence that primacy effects do occur regularly, and that questions in which respondents are asked to code a limited number of answers are particularly susceptible to them. We also have evidence that they occur in *code one only* questions and can occur in *code all that apply* questions. However for the latter question we also have evidence that recency effects can occur.

From these findings that primacy effects often occur we cannot, of course, infer that they lead to substantial response error. To asses this we need to assess their magnitude.

The magnitude of primacy effects

Table 2 shows the mean effect sizes for items presented first in each presentation order broken down by type of response instruction. Several findings are apparent in this table.

1. Taking all questions together there is evidence of primacy effects. On average the percentage of respondents picking an item when it was presented at the top of the list was over 1.6 percentage points greater than when it was presented at the bottom, irrespective of whether the item appeared at the top of the list in normal or reverse order implementation.
2. Effect sizes were highest when respondents were asked to pick a limited number of items from the list (*code n items* questions). For these types of question the percentage of respondents picking an item was around four percentage points greater when it was presented at the top of the list than when it was presented at the bottom.
3. For *code one only* questions, items presented at the top of the list were also, on average, chosen by significantly more respondents than were the same items when presented at the bottom of the list.
4. For *code all that apply* questions, items presented at the top of the list during reverse order implementation were, on average, chosen by significantly more respondents than were the same items when presented at the bottom of the list. More respondents chose items presented at the top of the list in normal order presentation than chose them when they were presented at the bottom of this list, although the difference was only marginally significant ($p<0.1$).

The findings of substantial effect sizes for *code n items* questions and lesser ones for *code one only* questions are largely in line with the findings on frequency of primacy effects discussed above.

For *code all that apply* questions the analyses of frequency and effect sizes are also congruent for items presented first in reverse order (both show evidence of small but significant primacy effects). However, for items presented first in normal order the two sets of findings appear to be in conflict: the analysis of frequency points to the presence of recency effects whereas the analysis of effect size shows marginally significant ($p<0.1$) primacy effects. The apparent conflict arises because, although primacy effects are less likely to occur than are recency effects, when they do occur they are of considerably greater magnitude (for items presented first in normal order presentation of *code all that apply* question the average recency effect was 1.36 percentage points average, whereas the average primacy effect was 3.77 percentage points).

Table 2 Effect size by question response coding instructions

	Mean % when presented first	Mean % when presented last	Mean effect size (mean % difference between when presented first and last)	Std. Deviation	Std. Error Mean	t	df	P under null hypothesis (1-tailed)
Normal order								
All questions	34.94	33.26	1.69	4.12	0.34	4.97	146	<0.01
Code one only	28.70	27.04	1.66	4.26	0.69	2.40	37	<0.05
Code all that apply	35.42	34.80	0.62	3.35	0.38	1.62	77	NS (<0.1)
Code n items	41.42	36.99	4.43	4.54	0.82	5.43	30	<0.01
Reverse order **All questions**	18.44	16.81	1.62	3.38	0.28	5.83	146	<0.01
Code one only	12.62	11.57	1.04	2.62	0.43	2.46	37	<0.01
Code all that apply	19.66	18.63	1.02	3.13	0.35	2.88	77	<0.01
Code n items	22.50	18.65	3.85	3.91	0.70	5.48	30	<0.01

The predictors of primacy effects

The analyses presented above established that both frequency and magnitude of primacy effects vary by question response instructions. We now examine whether a number of other question features predict primacy effects. As stated above, satisficing theory predicts that primacy effects will be more likely to occur when questions are harder to answer and when respondents are less motivated to answer them.

As stated above, from the literature on satisficing, we might expect to find:

- that susceptibility to primacy effects varies by response coding instructions (greatest with *code n items* questions and least with *code all that apply* questions);
- that susceptibility to primacy effects is greater for questions with more answer codes;
- that susceptibility to primacy effects is greater for questions with more words in the question stem;
- that susceptibility to primacy effects is greater for questions answered by large numbers of respondents;
- that susceptibility to primacy effects is greater for questions presented late in the questionnaire.

Logistic regressions were conducted separately for normal and reverse presentation orders using whether or not a primacy effect was observed as the dependent variable[4] and the following as independent variables:

- question response instructions (with *code all that apply* used as reference category);
- number of substantive codes on the showcard;

[4] coded 1 if primacy effect observed and zero if one was not observed.

- number of words in the question stem;
- whether the question was in first or second half of the questionnaire (with first half used as reference category);
- percentage of respondents picking the code.

These logistic regressions are summarised in tables 3 and 4.

Similarly, OLS regressions using the same independent variables and effect size as the dependent variable were run. The results of these are shown in tables 5 and 6.

Table 3 Logistic regression normal order

	B	SE	Sig	Odds ratio
Code n items	2.48	0.68	<0.001	11.91
Code one only	1.29	0.53	<0.05	3.62
Number of responses	-0.05	0.04	NS	0.95
Number of words in question stem	0.01	0.02	NS	1.01
Questionnaire location	0.68	0.41	<0.10	1.98
Mean % picking code	0.00	0.01	NS	1.00
Constant	-1.19	0.89	NS	0.30

Table 4 Logistic regression reverse order

	B	SE	Sig	Odds ratio
Code n items	1.67	0.62	<0.01	5.32
Code one only	0.80	0.54	NS	2.23
Number of responses	0.09	0.04	<0.05	1.09
Number of words in question stem	-0.02	0.02	NS	0.98
Questionnaire location	0.69	0.40	<0.10	2.00
Mean % picking code	0.02	0.01	NS	1.02
Constant	-1.51	0.89	<0.10	0.22

Table 5 OLS regression normal order (R2 = 0.18)

	B	SE	sig
Code n items	3.37	0.87	<0.001
Code one only	0.61	0.92	NS
Number of responses	-0.08	0.06	NS
Number of words in question stem	0.01	0.03	NS
Questionnaire location	1.56	0.66	<0.05
Mean % picking code	0.00	0.01	NS
Constant	-1.14	1.49	NS

Table 6 OLS regression reverse order (R2 = 0.20)

	B	SE	sig
Code n items	2.89	0.70	<0.001
Code one only	0.61	0.76	NS
Number of responses	0.02	0.05	NS
Number of words in question stem	-0.02	0.02	NS
Questionnaire location	0.05	0.53	NS
Mean % picking code	0.06	0.02	<0.001
Constant	-0.03	1.21	NS

For items presented first in the normal order the logistic regression showed that presence of primacy effects was significantly associated only with question response instructions (they were significantly more likely to be found for *code n items* questions and for *code one only* questions relative to *code all that apply* questions). Position in the questionnaire also was marginally significantly related to presence of primacy effects: primacy effects were more likely to be observed for questions presented in the second half of the questionnaire. The results of the OLS regression for primacy effect magnitude largely mirrored these findings: significant effects were found for question response instructions (although the contrast between *code one only* questions and *code all that apply* questions was no longer significant) and position in the questionnaire. There was no evidence of an association between presence or magnitude of primacy effects on the one hand and number of responses, number of words in the question stem, or proportion of respondents picking the item on the other.

For items presented first in the reverse order the logistic regression showed that the presence of primacy effects was significantly associated with question response instructions (significantly more likely to be found for *code n items* questions relative to *code all that apply* questions), number of responses (more primacy effects with longer lists). It was also marginally significantly related to position in the questionnaire (more effects for questions presented in the second half of the questionnaire). The results of the OLS regression for primacy effect magnitude showed a somewhat different pattern. Although significant effects were found for question response instructions (for the contrast between *code n items* questions and *code all that apply* questions), none were found for number of response items and position in the questionnaire. A significant positive relationship *was* found for proportion of respondents picking the item in the predicted direction. Again no relationship was found between either presence or magnitude of primacy effects and number of words in the question stem.

Discussion and conclusions

There are two main conclusions which the survey practitioner can draw from these analyses. First, primacy effects frequently occurred in the Ipsos MORI questions examined, although on average, they were small in magnitude.

Second, although evidence of primacy effects was found for the three question types examined, they were considerably more frequent and larger in magnitude for questions where respondents were asked to select a set number, or maximum number, of items where this number was greater than one (the *code n items* format). Of course, this finding should be interpreted with caution. This was not an experimental study, and the observed differences may have been caused by other differences in content or format between *code n items* questions and other types of question. That said, examination of the individual

questions did not throw up any obvious candidates for confounding variables, and we feel that it is advisable for the survey practitioner to use *code n items* questions with some caution until definitive experimental studies have been undertaken. Experiments which manipulate response coding instructions whilst holding all remaining aspects of question content constant should be simple to implement, and Ipsos MORI plans to undertake an initial investigation in the near future.

The predictions we made on the basis of the theory of satisficing received scant support. There was some slight evidence that questionnaire location may be related to a question's susceptibility to primacy effects, but other predictions relating to task difficulty and item plausibility did not receive meaningful support from these data. However, we acknowledge that our operationalizations of task difficulty and plausibility were crude and quite possibly inadequate to the task of testing the relevant hypotheses. This research was opportunistic in nature and we were, perforce, limited to available variables.

References

Duffy, R. (2003). Response order effects – how do people read? *International Journal of Market Research, 45*, pp457-466.

Knäuper, B. (1999a). Age differences in question and response order effects. In N. Schwarz, D. Park, B. Knäuper & S. Sudman (Eds.) *Cognition, aging, and selfreports* (pp.341–363). Philadelphia: Psychology Press.

Knäuper, B. (1999b).The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly, 63*, 347–370.

Krosnick, J. A. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys, in *Applied Cognitive Psychology*, *5*, pp 213-236, p.213.

Krosnick, J. A. and Alwin D.F. (1987) An evaluation of a cognitive theory of response-order effect in survey measurement, in *Public Opinion Quarterly*, *51*, pp 201-219, 1987.

Narayan, S. and Krsnick, J. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, 60*, 58-88.

Payne, S. (1951) *The Art of Asking Questions*. Princeton University Press.

Rugg, D. and Cantril, H. (1944) The wording of questions. In H. Cantril (ed.), *Gauging Public Opinion.* Princeton University Press

Schuman, H. and Presser, S. (1981) *Questions and answers in attitude surveys: Experiments in question form, wording and context*. Academic Press

Tourangeau R. , Rips L., and Rasinski K. (2000) *The Psychology of Survey Response* (Cambridge University Press).

Appendix: examples of question types

1. Code all that apply

SHOW CARD Which, if any, of the following do you think a dermatologist does? Just read out the letter or letters that apply.

A Treat acne
B Treat eczema
C Treat skin cancer
D Facelifts
E General health checks
F Treat hair loss
G Facials
H Mole checks
I Treatment/Diagnosis of rashes
J Tanning
K Waxing
None of these
Don't know

2. Code one only

SHOW CARD Thinking back to when you were at primary school (up to the age of 11 or 12), which of the following job or occupation categories most closely matches what you really wanted or hoped to do when you left school? We're interested in knowing what, back then, was your ideal or dream job or occupation. Please just read out the letter that applies.

A Something very senior
B Something professional which you need a degree (or equivalent level professional qualification) to do
C Something else professional
D Something administrative or secretarial
E Skilled tradesman/woman
F Something involving personal service
G Something involving sales or other customer service
H Something involving processing, manufacturing or machine operating
I Something creative, artistic or sporting
J Something which you don't need any qualifications to do
Other (WRITE IN)
Did not have an ideal or dream job/occupation
Don't know

3. Code n items

SHOW CARD Thinking about buying chicken that is fresh or frozen, which of the following, if any, are of concern to you? PROBE FULLY - And any others? CODE UP TO A MAXIMUM OF THREE

A) Chicken being fed genetically modified foods
B) Use of growth hormones
C) Country of origin / possibility that chicken is imported
D) Unclear labelling on packaging
E) Animal welfare / conditions of rearing
F) Added water
G) Use of antibiotics
None of these
Don't shop / eat chicken
Don't know