Eliminating order effects in association tasks without using randomisation

Ian Durbach
University of Cape Town, African Institute for Mathematical Sciences, and Ipsos Laboratories
Gareth Lloyd
Ipsos Laboratories

It has often been observed that changing an item's position in a list can substantially affect the probability that it is chosen. This paper assesses the magnitude of these so-called order effects in brand-attribute association tasks, and examines the confounding roles played by brand usage and question framing. While our main order effect is roughly the same as that observed for similar response formats, we find substantially larger order effects among users of a brand than non-users; and question frames that first ask respondents to create an attribute shortlist before making associations on this reduced set eliminate or greatly reduce the magnitude of the order effect and its interaction with brand usage. These simple modifications to question framings may be useful where randomisation is not feasible.

Introduction

Brand-attribute association questions are ubiquitous in commercial marketing research. These typically require respondents to indicate which brands they believe possess certain properties or attributes, with the aim of establishing which brand characteristics are important to people in a market and measuring perceptions of brand performance on these characteristics. Strategy then involves maintaining or increasing brand performance on important attributes or identifying brand strengths that may not be important at the moment, but that could be leveraged if their importance could be increased.

Received (in revised form): 15 January 2014

DOI: 10.2501/IJMR-2015-062

Naturally there are many ways of asking the general association questions. Typically (in commercial research at least) respondents are shown an $I \times J$ matrix, where entry a_{ij} contains the association between brand $i \in \{1, 2, ..., I\}$ and attribute $j \in \{1, 2, ..., J\}$, and are asked to tick cells where they believe an association exists. But many other formats are possible. For example, should respondents be asked to respond on binary, ordinal or cardinal scales? Should brand and/or attribute lists be pre-specified or open-ended? Should limits be imposed on the number of associations given (overall, or for any one brand and/or attribute)? The vast literature on framing effects (e.g. Levin *et al.* 1998; McKenzie & Nelson 2003) suggests that these choices will have a substantive effect on the obtained responses.

One framing effect that has received special attention is the effect that the position of an attribute in the list has on its propensity to be picked – commonly referred to as position bias or the order effect. A large number of studies have examined the nature and size of order effects in various types of survey responses (Day 1969; Jain & Pinson 1976; Chan 1991; Winchester *et al.* 2008). These studies overwhelmingly demonstrate that a respondent's choice of an answer from a list of alternatives depends on the position of the answer in relation to other alternatives, as well as on the content of that answer. While we have been unable to find any studies specifically examining order effects in attribute association lists, many of the findings obtained using other formats are directly relevant.

The current paper assesses the magnitude of order effects in attribute association questions, with a special emphasis on examining the interaction between attribute order effects and the way in which the association question is framed. We do this using a survey design that randomly assigns respondents to a framing format and randomises the order in which attributes are presented to the respondent. Eight framings are used. We first allocate respondents to either an 8-attribute or 18-attribute treatment, and then to one of four possible presentation styles: a standard attribute association matrix with or without limitations on the number of associations that may be made; and two frames in which subjects first create an attribute shortlist before making associations on this reduced set.

The primary objective of the paper is to evaluate whether any framing approach is able to reduce the magnitude of the order effect. We are unaware of any previous comparative evaluation based on alternate formats for attribute association questions. The presence of an interaction between question framing and the order of attribute presentation is of particular interest for those conducting market research in developing countries, where pen-and-paper surveys are still commonplace and complete randomisation is not always feasible. In these contexts order effects have a direct negative effect on the robustness of research findings, calling into question the conclusions drawn from the data and subsequently reported to clients. Using more effective question framings, where possible, has a direct bearing on the value of research.

The remainder of the paper is structured as follows. The next section gives a brief overview of the literature on order and framing effects. We then describe our research methodology: data collection, sampling details and research hypotheses. The section after that provides results and a discussion of the practical implications for researchers. A final section concludes the paper and provides some recommendations for practitioners.

Literature review

'Framing effects' occur when different descriptions of a problem lead to different responses to that problem. A large body of research has shown that much of human judgement and decision making is affected by framing effects (see, for example, Tversky & Kahneman 1981; Levin *et al.* 1998; McKenzie & Nelson 2003). Of course, differences in response patterns in the face of changes to question wording, response options, visual design, and so on, are highly relevant to survey design, and have been extremely well documented in, for example, Bradburn (1982), Sudman and Bradburn (1982), Converse and Presser (1986), Dillman *et al.* (2009) and Bruine de Bruin (2011). One classic survey-related framing effect is the so-called 'order effect', first quantified by Payne (1951), who reported that the response frequency of an answer option increased by, on average, six (two) percentage points when placed among the top (bottom) alternatives, relative to when it was placed near the middle of the list. Since then much research has focused on the role of response format and respondent and interviewer characteristics on the presence and magnitude of the effect.

Order effects have been reported across a spectrum of response formats: multiple-choice questions (Blunch 1984), pairwise comparisons (Day 1969), multidimensional scales (Jain 1976), paired product tests (Day 1969), negative image attributes (Winchester *et al.* 2008) and Likert scales (Chan 1991). A consistent finding is that attributes are more likely to be selected when they appear early on in written response formats (the 'primacy' effect) but later on when the response format is verbal (the 'recency' effect). The most convincing explanation for these effects is that they are manifestations of 'survey satisficing' (e.g. Krosnick 1991), in which respondents use a number of cognitive heuristics to 'minimise effort in responding to surveys', while 'providing the appearance of compliance' (Malhotra 2009). Two cognitive mechanisms used by the general satisficing strategy interact to produce order effects (Holbrook *et al.* 2007). These are confirmation bias, a tendency to seek reasons to select a particular option rather than reasons why not to select it (Koriat *et al.* 1980; Klayman & Ha 1987; Nickerson 1998), and a tendency to stop evaluating options once enough evidence has been given (i.e. attributes have been ticked) to justify why a particular brand is used or liked. In general, order effects arise because, as a respondent's reserve of available effort and motivation are depleted, response strategies shift from 'optimal' to heuristic, including less effortful searches of memory (i.e. for reasons justifying the selection of a response) and early stopping of the response task.

Early attempts to address order effects focused on the randomisation of the positions that responses occupy within a list (Payne 1951; Green & Tull 1978), and indeed this common-sense response is widely used in practice today. Blunch (1984), however, showed that randomisation will in most cases not remove position bias in its statistical sense (i.e. the difference between an estimator and the population parameter it is meant to estimate), so that we cannot say that bias is absent when an alternative receives the same support regardless of its position – although this may be true, we can only say with certainty that the various positions are equally biased. A related point is that randomisation does not allow for estimation of the size of the position bias.

Research methodology

Data

Our data consist of an online survey of 4,000 study participants recruited from a US-based web panel managed by legacy Synovate Viewsnet (now Ipsos) and conducted in the laundry detergents category. All respondents were 18 years or over and aware of at least one brand of laundry detergent. The questionnaire included two attribute association tasks to measure brand image. The first consisted of 8 attributes, the second 18. The attributes used were as follows.

- 8-attribute list: Keeps my family looking good; Helps me feel confident with my appearance; Is currently a leading brand; Is a family favourite my mother has always used; Is a fun brand; Cares about the environment; I would be proud to use; Gives me value for money.

- 18-attribute list: Provides superior whiteness; Keeps colours bright; Leaves clothes smelling fresh; Is good for a variety of fabrics; Is an innovative brand; Is a reliable laundry detergent; Is a laundry detergent that works well; Has a wide variety of scents; A brand that makes my clothes soft; A brand that gives me more washes; Has a long-lasting scent; A brand that is gentle on my clothes; Is safe for sensitive skin; Is an environmentally friendly brand; Gets clothes really clean; Has a scent I like; A brand that is value for money; My kind of brand.

Respondents were asked to rate all brands of which they were aware. On average, respondents were aware of 10.1 brands (from a list of 16). The sample of 4,000 was then distributed equally between the four different presentation formats. The basic presentation layout remained the same (brands across the top, attributes down the left-hand side) for all formats, which conforms to the standard association matrix grid commonly used in research surveys. The difference between the four cells was how respondents were asked to answer the question. The formulations were as follows.

- Standard: a standard brand-by-attribute association matrix.
- Limited: the same association matrix but with respondents asked to pick at least two but not more than five for each brand that's relevant to them.
- Shortlist v1: respondents are first asked to identify 'attributes they cannot do without' and then were shown a reduced brand-by-attribute matrix including only those attributes identified in the first step.
- Shortlist v2: respondents are first asked to identify their 'most important attributes' and then were shown a reduced brand-by-attribute matrix including only those attributes identified in the first step.

For each response task, the order in which attributes were presented was randomised.

Hypotheses

Order effects are primarily driven by heuristic response strategies employed when time, interest, or cognitive ability and effort is limited or impaired. Theories of survey satisficing (Krosnick 2000) suggest that respondents have limited supplies of motivation and effort that they are willing to expend during response tasks. If this reserve is depleted before the task is complete, respondents will either stop the process of answering or change their response strategies to ones that are less cognitively demanding, including insufficient or biased searches of memory for relevant information and/or insufficient or biased integration of information into a summary judgement. Order effects are one manifestation of survey satisficing. Of course, we expect an order effect to be present in attribute association tasks (H1a), since these are similar in essence to many of the other judgement tasks in which order effects have been previously observed.

Further to this, survey satisficing explicitly predicts that, once effort supplies have been depleted and 'questionnaire completion continues, respondent motivation may decrease and fatigue may increase, leading to *further* degradation in the response process' (Vannette & Krosnick 2014, emphasis added). Thus longer attribute lists are expected to elicit greater use of heuristic strategies and hence exhibit larger order effects (H1b). Note that, because we have entirely different attribute lists in the 8-attribute and 18-attribute conditions, the effect of attribute number is confounded with the effect of which attributes were selected to appear on each list. However, there appear to be no substantive differences between the broad types of attribute populating the two lists.

Then, we expect confirmatory biases, and hence order effects, to be stronger for used brands than non-used brands, since the fact of usage provides a strong cue for selecting a positive attribute (which all attributes in our study are). We thus hypothesise that order effects will be more pronounced for those brands that a respondent uses regularly, relative to those of which he or she is merely aware (H2).

Framing features that reduce the cognitive load on the respondent are hypothesised to reduce both the magnitude of the mean order effect (H3a) and the interaction between attribute position and brand usage (H3c). While it is not indisputable that the shortlist formats reduce cognitive load (since they replace a single association task with an importance-assessment task and an association task) we argue that, because of the substantial reduction in the number of judgements that must be made (at least in principle), such a reduction is indeed likely. Note that the shortlist formats replace one set of $I \times J$ binary association judgements, where I is the number of brands and J is the number of attributes, with an initial set of J binary importance judgements and secondary set of $I \times J_S$ binary association judgements, where J_S is the number of shortlisted attributes, i.e. a total of $J + (I \times J_S)$ judgements. For our data (see below) the number of shortlisted attributes J_S was on average half of the full list size J, so that the alternate formats require on average $IJ - J(1 + 0.5I) = J(0.5I - 1)$ fewer judgements. Even for moderate sized tasks such as ours, where $J = 8(18)$ and $I = 16$, this equates to a substantial reduction of 56 (126) judgements (out of an initial total of 128 (288)).

In essence, we hypothesise that the alternate formats will to some extent counteract the tendencies outlined in H1a and H2. Since the preliminary shortlisting of attributes potentially removes a great deal of information for the respondent to consider, we hypothesise a greater reduction in the main order and interaction effects for those presentation formats relative to one in which the size of the problem remains the same and only the number of associations is limited (H3b and H3d, respectively). Our seven hypotheses are summarised below.

H1a: The position of an attribute in a brand association list affects the likelihood of it being picked.

H1b: Longer attribute lists will show larger order effects than shorter lists.

H2: Larger order effects will be observed for used than non-used brands.

H3a: Framing formats that reduce the number of associations to be made will counteract H1a, i.e. will reduce the magnitude of the mean order effect.

H3b: Framing modes that explicitly create a shortlist of attributes will be more effective in reducing the size of the basic order effect.

H3c: Framing formats that reduce the number of associations to be made will counteract H2, i.e. will reduce the interaction between attribute position and brand usage.

H3d: Framing modes that explicitly create a shortlist of attributes will be more effective in reducing the size of the interaction between attribute position and brand usage.

Statistical analysis

To each of the 8- and 18-attribute datasets we fit a generalised linear mixed model (GLMM) to the binary response that subject k selects an attribute j for brand i, with two random intercepts $v_k \sim N(0, \sigma_v^2)$ and $w_j \sim N(0, \sigma_w^2)$ accounting for between-subject and between-attribute heterogeneity, respectively. Independent variables are the rank in the list where the attribute appears (r_{ijk}), the frame used (indicator variables f_{2ijk}, f_{3ijk}, f_{4ijk} for the Limited, Shortlist v1 and Shortlist v2 frames respectively), and whether the brand is used or not (u_{ijk}). Testing the hypotheses above requires that we fit all two- and three-way interactions, i.e.

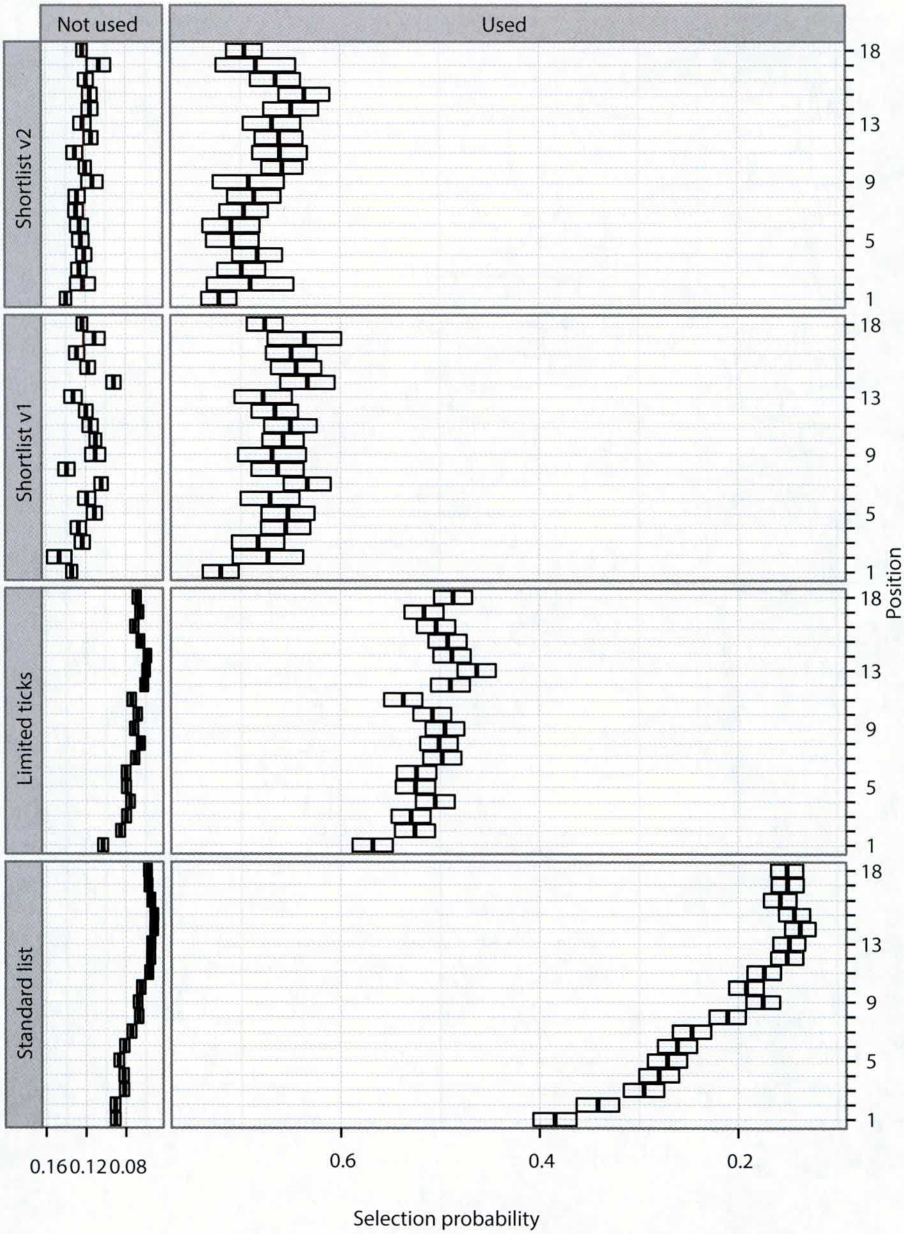
$$\begin{aligned}\text{logit}(p_{ijk}) = {} & \beta_0 + \beta_1 r_{ijk} + \beta_2 u_{ijk} + \beta_3 f_{2ijk} + \beta_4 f_{3ijk} + \beta_5 f_{4ijk} + \beta_6 r_{ijk} u_{ijk} \\ & + \beta_7 r_{ijk} f_{2ijk} + \beta_8 r_{ijk} f_{3ijk} + \beta_9 r_{ijk} f_{4ijk} + \beta_{10} u_{ijk} f_{2ijk} + \beta_{11} u_{ijk} f_{3ijk} \\ & + \beta_{12} u_{ijk} f_{4ijk} + \beta_{13} r_{ijk} u_{ijk} f_{2ijk} + \beta_{14} r_{ijk} u_{ijk} f_{3ijk} + \beta_{15} r_{ijk} u_{ijk} f_{4ijk} \\ & + v_k + w_j\end{aligned}$$

where $p_{ijk} = \Pr(a_{ijk} = 1)$.

Results

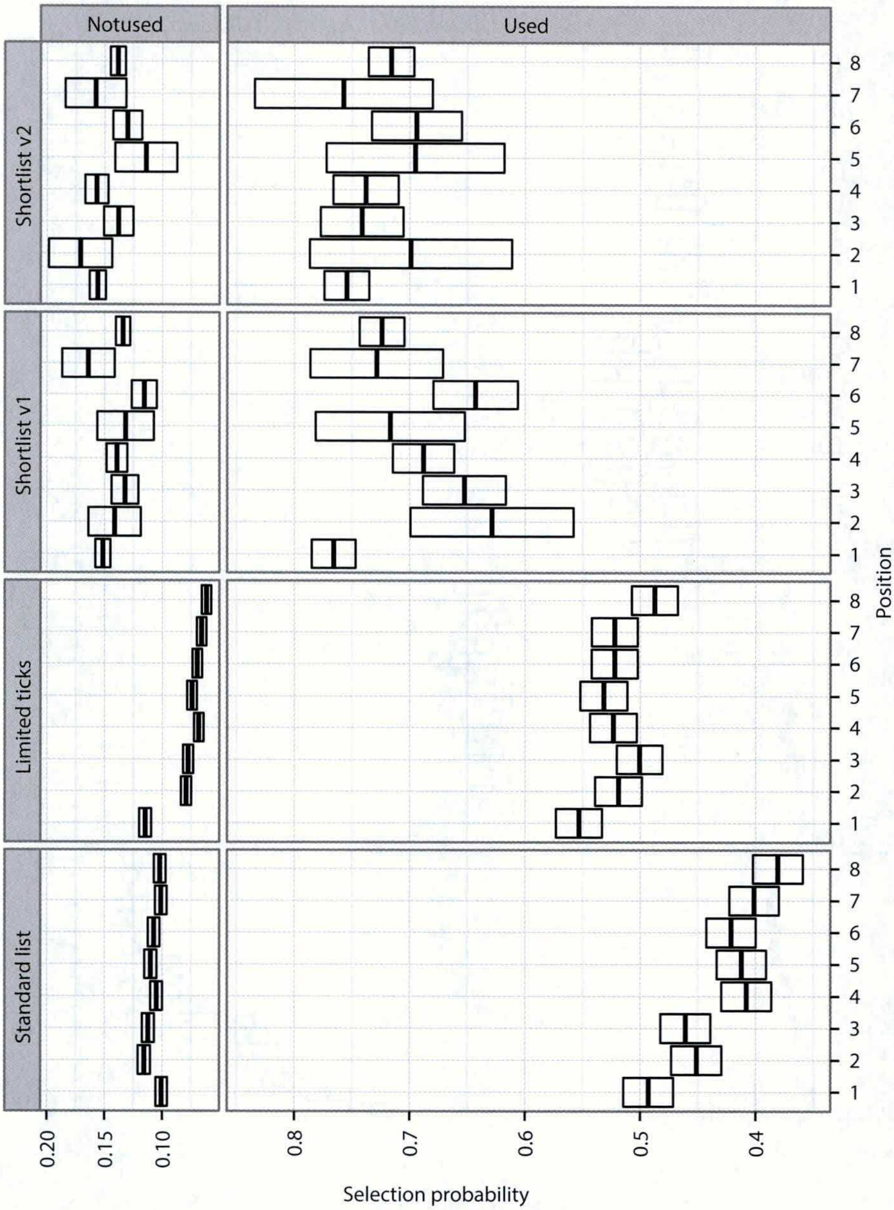
Before presenting our main model results, Figures 1 and 2 show the basic pattern of the order effect in the four framing conditions, for users and aware non-users, and for 18- and 8-attribute cases, respectively. Note that, in the two framings employing preliminary shortlists, most respondents will select considerably fewer than the full set of 18 (or 8) attributes. In the 18-attribute case, subjects selected an average of 10.7 attributes; in the 8-attribute case, they selected on average 3.8. To make the four framing conditions comparable, we transformed each attribute's position in the shortlist into an equivalent rank in the full 18- or 8-attribute list. For example, if a respondent identified 10 out of 18 attributes as important, the attribute appearing in rank 6 of the shortlist of 10 has a transformed rank of 6/10 x 18 = 10.8, which is rounded to 11.

A number of preliminary observations can be made from these figures. First, while a clear order effect exists in the standard 18-attribute list for both users and aware non-users, selection probabilities in the shorter 8-attribute list show a much smaller influence of attribute position, with little or no effect being observed for non-users. This suggests that order effects are certainly possible in attribute association tasks as traditionally carried out in commercial marketing research practice, but that the length



Note: The central line of each box denotes the mean, with the upper and lower boundaries of the box indicating 95% upper and lower confidence limits for that mean

Figure 1 Mean selection probability in each position of the 18-attribute list, for each of the four framing conditions used (i.e. averaged over all brands, attributes and subjects), and for aware non-users and users of a brand, respectively



Note: The central line of each box denotes the mean, with the upper and lower boundaries of the box indicating 95% upper and lower confidence limits for that mean

Figure 2 Mean selection probability in each position of the 8-attribute list, for each of the four framing conditions used (i.e. averaged over all brands, attributes and subjects), and for aware non-users and users of a brand, respectively of the attribute list is a key consideration. In practice, most association tasks involve more than 20 attributes (in our experience at least), so that order effects are likely to be ubiquitous in commercial market research. The possibility of reducing these effects should provide a strong incentive to keep attribute lists as short as possible, however.

In the 18-attribute case, the difference between the maximum selection probability at the top of the list and the minimum close to but not at the bottom is 24% for users and 6% for non-users. The latter figure is very similar to what has been observed in other studies, but for users the size of the effect is considerably larger. Equivalent effect sizes in the 8-attribute case are 10% (users) and 0% (non-users). This of course leaves the relationship between list size of the magnitude of the order effect unspecified: with only two conditions we cannot assess this in any detail. What is clear, though, is that there is no single 'size' of order effect in attribute association tasks – this will be highly contingent on the size of the attribute list, and quite probably on other factors, too.

Figure 1 also shows evidence that alternate framings, particularly those that ask respondents to form a preliminary shortlist of attributes, are able to substantially reduce the size of the order effects. The greater variability around the mean in the two shortlist framings is primarily due to smaller sample sizes (because of the reduced attribute lists constructed by respondents) rather than more variable behaviour. The effect of the alternate framings is more difficult to discern in Figure 2, because order effects are smaller to begin with, but in both cases there is evidence to suggest that the alternate framings may be helpful in reducing the size of the order effect.

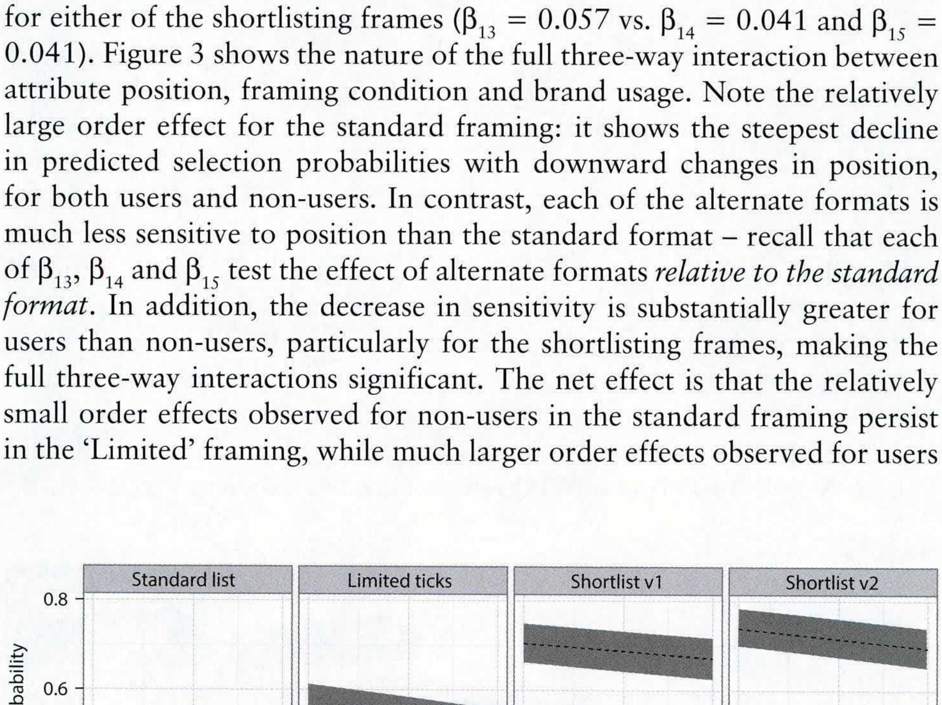
Table 1 shows parameters and significance values (z-statistics) for our fitted models. We first discuss results for the 18-attribute list before turning to the more problematic 8-attribute case. The negative coefficient associated with an attribute's position r_{ijk} confirms the existence of a significant order effect using the standard presentation format, even where a linear relationship has been imposed. However, if an alternate format is used, the size of this effect can be greatly reduced. Note that β_7 is positive and roughly half the magnitude of the basic order effect captured by β_1, while β_8 and β_9 (which model the effect of the two shortlisting frames) are roughly 85% of the magnitude of β_1. Thus, while the alternate formats do not remove the order effect entirely, they significantly counteract it – as hypothesised by H3a. H3b, which hypothesised a greater effect for the shortlist-creating frames, is also confirmed.

Table 1 Parameters and test statistics for a generalised linear mixed model fitted to the probability that an attribute is selected, using attribute position, brand usage and framing mode as independent variables

					18-attributes			8-attributes		
Effect	**Description**	**Variable**		H_0	**Estimate**	*SE*	*z*	**Estimate**	*SE*	*z*
β_0	Intercept				−1.879	0.103	−18.3	−1.707	0.134	−12.7
β_1	Rank	r_{ijk}	H1a	<0	−0.045	0.002	−26.9	−0.008[b]	0.005	−1.8
β_2	Used	u_{ijk}			1.209	0.032	37.5	1.655	0.046	36.3
β_3	Limited format	f_{2ijk}			−0.259	0.051	−5.1	−0.140	0.054	−2.6
β_4	Shortlist v1 format	f_{3ijk}			0.080	0.053[b]	1.5	0.307	0.060	5.2
β_5	Shortlist v2 format	f_{4ijk}			0.114	0.053[a]	2.2	0.348	0.062	5.6
β_6	Rank x Usage	$r_{ijk}u_{ijk}$	H2a	<0	−0.044	0.003	−13.7	−0.056	0.009	−6.2
β_7	Rank x Limited	$r_{ijk}f_{2ijk}$	H3a	>0	0.019	0.002	8.1	−0.079	0.007	−10.9
β_8	Rank x Shortlist v1	$r_{ijk}f_{3ijk}$	H3b	$>\beta_7$	0.038	0.003	15.2	−0.017[a]	0.008	−2.1
β_9	Rank x Shortlist v2	$r_{ijk}f_{4ijk}$	H3b	$>\beta_7$	0.035	0.003	13.7	−0.017[a]	0.008	−2.1
β_{10}	Used x Limited	$u_{ijk}f_{2ijk}$			1.178	0.045	26.0	0.440	0.067	6.6
β_{11}	Used x Shortlist v1	$u_{ijk}f_{3ijk}$			1.471	0.053	27.8	0.871	0.085	10.3
β_{12}	Used x Shortlist v2	$u_{ijk}f_{4ijk}$			1.615	0.054	29.6	0.965	0.091	10.7
β_{13}	Rank x Used x Limited	$r_{ijk}u_{ijk}f_{2ijk}$	H3c	>0	0.057	0.004	13.1	0.122	0.013	9.3
β_{14}	Rank x Used x Shortlist v1	$r_{ijk}u_{ijk}f_{3ijk}$	H3d	$>\beta_{13}$	0.041	0.005	8.4	0.059	0.016	3.7
β_{15}	Rank x Used x Shortlist v2	$r_{ijk}u_{ijk}f_{4ijk}$	H3d	$>\beta_{13}$	0.041	0.005	8.2	0.050	0.017	3.0
Effect	**Description**	**Variable**			**Estimate**	*SD*		**Estimate**	*SD*	
σ_v^2	Respondent-level var.	v_k			0.991	0.996		0.798	0.893	
σ_w^2	Attribute-level var.	w_j			0.168	0.409		0.133	0.364	

Notes: The last two rows show results for the random component of the model, in which random intercepts are fitted for each respondent and attribute; superscript *a* and *b* denote effects not significant at 1% and 5% level, respectively

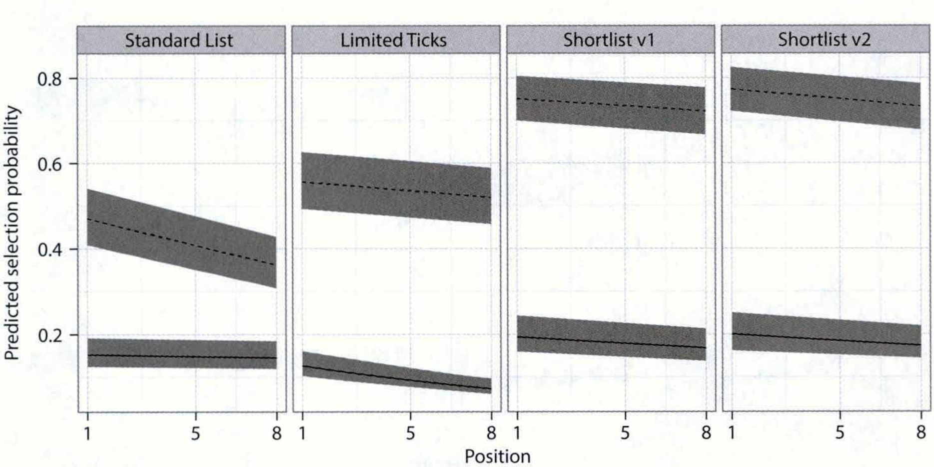
The highly significant negative coefficient $\beta_6 = -0.044$ confirms that order effects are larger for used brands than those of which the respondent is merely aware (H2). This phenomenon, however, is eliminated if an alternate format is used (H3a: see the positive values for β_{13}, β_{14} and β_{15} equal or greater in the magnitude of the basic interaction effect captured by β_6). However H3b, which hypothesised a greater effect for the shortlist-creating frames, is not confirmed. In fact, the opposite is true, as shown by the significantly larger effect size for the 'Limited' frame than for either of the shortlisting frames ($\beta_{13} = 0.057$ vs. $\beta_{14} = 0.041$ and $\beta_{15} = 0.041$). Figure 3 shows the nature of the full three-way interaction between attribute position, framing condition and brand usage. Note the relatively large order effect for the standard framing: it shows the steepest decline in predicted selection probabilities with downward changes in position, for both users and non-users. In contrast, each of the alternate formats is much less sensitive to position than the standard format – recall that each of β_{13}, β_{14} and β_{15} test the effect of alternate formats *relative to the standard format*. In addition, the decrease in sensitivity is substantially greater for users than non-users, particularly for the shortlisting frames, making the full three-way interactions significant. The net effect is that the relatively small order effects observed for non-users in the standard framing persist in the 'Limited' framing, while much larger order effects observed for users



Note: Solid and dashed lines show predicted mean selection probabilities for non-users and users respectively, as obtained from the GLMM described in the text; grey bands around the lines indicate upper and lower 95% confidence intervals

Figure 3 Graphical illustration of three-way interaction between attribute position, framing condition and brand usage for the 18-attribute case are much reduced. In contrast, the shortlisting frames eliminate or greatly reduce the order effect among both users and non-users.

In the 8-attribute case the magnitude of the basic order effect is smaller than in the longer list and is only weakly statistically significant at p = 0.07 (H1, β_1). The effect remains substantially and significantly larger among users than among aware non-users (H2, β_6), with the size of the usage effect being similar in magnitude to that observed in the longer 18-attribute list. H3a and H3b, which posit that the alternate formats will reduce the size of the basic order effect, are not supported, as the associated coefficients (β_7, β_8, β_9) are all negative. As suggested by Figure 2, the principal reason for this is the small order effect observed for aware non-users, who make up the majority of the sample. Among users however, all three alternate formats do indeed substantially reduce the magnitude of the order effect relative to the standard format (H3c and H3d: β_7, β_8, β_9). Note that the total per-unit order effect for users in the standard case is $\beta_1 + \beta_6 = -0.057$. The marginal contribution of the 'Limited' format is to reduce this order effect by $\beta_7 + \beta_{13} = 0.043$; for the two shortlisting formats the equivalent contributions are $\beta_8 + \beta_{14} = 0.042$ and $\beta_9 + \beta_{15} = 0.033$. That is, in each case the alternate formats act to substantially reduce the size of the order effect. Figure 4 shows the three-way interaction between attribute position, framing condition and brand usage. Finally, in both 8- and 18-attribute conditions between-respondent variability, σ_v^2



Note: Solid and dashed lines show predicted mean selection probabilities for non-users and users, respectively, as obtained from the GLMM described in the text; grey bands around the lines indicate upper and lower 95% confidence intervals

Figure 4 Graphical illustration of three-way interaction between attribute position, framing condition and brand usage for the 8-attribute case was observed to be substantially greater than attribute-level variability σ_w^2, although likelihood ratio tests indicated better model fit if both sources of heterogeneity were included.

Conclusions

The observation that changing an item's position in a list of alternatives can substantially affect the chance that it is selected is an enduring and well-studied feature of survey research. Respondents with limited time and motivation may, instead of devoting their full effort and attention to answering a survey question, fall back on simple heuristic strategies that allow the survey question to be adequately completed with only a reasonable expenditure of time and effort. Previous studies suggest that two mechanisms that influence this general strategy of survey satisficing are a confirmation bias, the tendency to seek reasons for choosing an option rather than not choosing it; and an early termination of the response process once a suitable alternative has been identified or enough information has been given to provide 'the appearance of compliance' (Malhotra 2009).

While order effects have not, to our knowledge, been directly examined in the brand-by-attribute binary association matrices that are so common in commercial market research, there is no reason why this particular format should be exempt from an effect that has been observed in so many related formats. Indeed, using a standard attribute association task involving a moderate-sized problem (from the perspective of commercial market research) consisting of 18 attributes and 16 brands, we showed that the overall size of the order effect, i.e. range of selection probabilities, was approximately 7%, similar to what has been observed in other studies (e.g. Payne 1951). We found, however, a highly significant interaction between brand usage and attribute position, such that the magnitude of the observed order effects was substantially larger among users of a brand than among those who were merely aware. Our conjectured explanation for this observation is that brand usage provides a strong cue for selecting a positive attribute, and thus for confirmatory biases to be stronger for used brands than non-used brands. We thus hypothesised that order effects will be more pronounced for those brands a respondent regularly uses, relative to those of which he or she is merely aware. Among brand users, we found order effects of the order of 25% – considerably larger than is typically reported in samples as a whole.

A typical response to the existence of order effects is to randomise the order in which attributes are presented. This, however, requires relatively larger sample sizes in order for the experimental design to remain balanced and in addition may not always be practically feasible – for example, where responses are collected using pen and paper, and field workers may lack the necessary tools to implement randomisation in the process of data collection. A main goal of the current paper has thus been to assess whether different ways of asking the attribute association question might lead to smaller order effects than the standard format. Since order effects are primarily results of resource constraints – time, ability and motivation – any format that simplifies the ensuing response process is expected to reduce the size of the effect. This is reflected in the fact that, even with the standard format, we found much smaller order effects using an 8-attribute list than if 18 attributes were included, although here it must be noted that because different attribute lists were used in the 8- and 18-attribute conditions, we cannot rule out the possibility of confounding effects.

We conducted an experiment using three alternate formats: one in which the number of ticks that each brand could receive was limited to be between two and five; and two in which respondents were asked to first identify which attributes were most important to them and then, in a second stage, were asked to provide associations using only this reduced attribute set. These two methods differed only in the benchmark set on how important an attribute needed to be to pass into the second stage: in one format we asked for attributes respondents 'could not do without', while in the other we simple asked for the 'most important' attributes.

In both attribute lists, we observed that the alternate formats eliminated or greatly reduced the magnitude of the order effect. In the shorter attribute list, a substantive order effect persisted for non-users if the less aggressive format limiting only the number of ticks was used. For both shortlisting formats, order effects were effectively eliminated among both users and aware non-users. Thus the simple mechanism of limiting the number of attributes that the respondent needs to consider when making the association, either directly (by using shorter attribute lists) or indirectly (by asking respondents to choose for themselves which attributes they want to enter into the association task), appears, to a large extent, to ameliorate the need for randomisation. Of course, where it is possible, randomisation remains the preferred course of action, since order effects can be assessed and explicitly accounted for. Shortened lists may reduce the size of the effect but, in the context of a single study, it will not be possible to assess whether the effect still exists and, if it does, whether it is substantial. This is an important practical point. However, in cases where randomisation is not possible, the shortlisting formats seem vastly preferable to the standard approach whenever the set of initial attributes is of a moderate or larger size. Our experience is that attribute lists in commercial market research typically contain at least 15 attributes – certainly at this size or larger, the reduced formats are strongly recommended.

References

Blunch, N. (1984) Position bias in multiple-choice questions. *Journal of Marketing Research*, **21**, pp. 216–220.

Bradburn, N.J. (1982) Question-wording effects in surveys, in Hogarth, R.M. (ed.) *Question Framing and Response Consistency*. Jossey-Bass, pp. 65–76. San Francisco: Jossey-Bass.

Bruine de Bruin, W. (2011) Framing effects in surveys: how respondents make sense of the questions we ask, in Keren, G. (ed.) *Perspectives on Framing*. Taylor & Francis, pp. 303–324. London: Taylor and Francis.

Chan, J.C. (1991) Response-order effects in Likert-type scales. *Educational and Psychological Measurement*, **51**, 3, pp. 531–540.

Converse, J.M. and Presser, S. (1986) *Survey Questions. Handcrafting the Standardized Questionnaire*. Thousand Oaks (California): Sage Publications.

Day, R.L. (1969) Position bias in paired product tests. *Journal of Marketing Research*, **6**, pp. 98–100.

Dillman, D.A., Smyth, J.D. & Christian, L.M. (2009) *Internet, Mail, and Mixed-mode Surveys. The Tailored Design Method*. New York: Wiley.

Green, P.E. & Tull, D.S. (1978) *Research for Marketing Decisions*, 4th edn. New Jersey: Prentice Hall.

Holbrook, A.L., Krosnick, J.A., Moore, D. & Tourangeau, R. (2007) Response order effects in dichotomous categorical questions presented orally: the impact of question and respondent attributes. *Public Opinion Quarterly*, 71, 3, pp. 325–348.

Jain, A.K. & Pinson, C. (1976) The effect on order of presentation of similarity judgements on multidimensional scaling results: an empirical examination. *Journal of Marketing Research*, **13**, pp. 435–439.

Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 2, pp. 211–228.

Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980) Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, pp. 107–118.

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

Krosnick, J.A. (2000) The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Newsletter* **20**, 1, pp. 4–8.

Levin, I.P., Schneider, S.L. & Gaeth, G.J. (1998) All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76, pp. 149–188.

Malhotra, N. (2009) Order effects in complex and simple tasks. *Public Opinion Quarterly*, 73, 1, pp. 180–198.

McKenzie, C.R. & Nelson, J.D. (2003) What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review*, **10**, 3, pp. 596–602.

Nickerson, R.S. (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, **2**, 2, pp. 175–220.
Payne, S.L. (1951) *The Art of Asking Questions*. New Jersey: Princeton University Press.
Sudman, S. and Bradburn, N.M. (1982) *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
Tversky, A. & Kahneman, D. (1981) The framing of decisions. *Science*, **211**, pp. 453–458.
Vannette, D. & Krosnick, J.A. (2014) *Answering Questions: A Comparison of Survey Satisficing and Mindlessness. The Wiley Blackwell Handbook of Mindfulness*, Vol. 1. John Wiley & Sons, pp. 312–327. New York: Wiley.
Winchester, M., Romaniuk, J. & Bogomolova, S. (2008) Positive and negative brand beliefs and brand defection/uptake. *European Journal of Marketing*, **42**, 5/6, pp. 553–570.

About the authors

Ian Durbach holds joint positions as Adjunct Senior Lecturer in the Department of Statistical Sciences at the University of Cape Town and a researcher at the African Institute for Mathematical Sciences. He has research interests in aspects of human decision making involving choice under uncertainty, applied statistical modelling and machine learning, mainly applied to marketing research and environmental and ecological modelling. He has acted as a Statistical Consultant to Ipsos Laboratories for several years.

Gareth Lloyd is a Research and Insights Director in the R&D division at Ipsos Laboratories with a specialization in brand-equity measurement. He is particularly interested in the role of behavioural economics and heuristics and biases in the consumer decision-making process. His research interests also include understanding how heuristics and biases affect survey responses and data outputs.

Address correspondence to: Ian Durbach, Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa.

Email: ian.durbach@uct.ac.za