Journal of Environmental Economics and Management 63 (2012) 73–91

Contents lists available at SciVerse ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Ordering effects and choice set awareness in repeat-response stated preference studies

Brett Day [a,*], Ian J. Bateman [a], Richard T. Carson [b], Diane Dupont [c], Jordan J. Louviere [d], Sanae Morimoto [e], Riccardo Scarpa [f], Paul Wang [d]

[a] University of East Anglia, Norwich NR4 7TJ, England, UK
[b] University of California, San Diego, USA
[c] Brock University, Canada
[d] University of Technology, Sydney, Australia
[e] Kobe University, Japan
[f] University of Waikato, New Zealand

ARTICLE INFO

Article history:
Received 17 June 2008
Available online 29 September 2011

Keywords:
Stated preference
Discrete choice experiment
Ordering effect
Advanced disclosure

ABSTRACT

We present an experiment designed to investigate the presence and nature of ordering effects within repeat-response stated preference (SP) studies. Our experiment takes the form of a large sample, full-factorial, discrete choice SP exercise investigating preferences for tap water quality improvements. Our study simultaneously investigates a variety of different forms of position-dependent and precedent-dependent ordering effect in preferences for attributes and options and in response randomness. We also examine whether advanced disclosure of the choice tasks impacts on the probability of exhibiting ordering effects of those different types. We analyze our data both non-parametrically and parametrically and find robust evidence for ordering effects. We also find that the patterns of order effect in respondents' preferences are significantly changed but not eradicated by the advanced disclosure of choice tasks a finding that offers insights into the choice behaviors underpinning order effects.

1. Introduction

A central tension in the advancement of stated preference (SP) methods of non-market valuation concerns the possible trade-offs between the desire to increase the information recovered from each survey respondent and the need for that additional information to usefully inform on respondents' preferences. With regards to information recovery, one clear trend has been the development of formats that present each respondent with multiple valuation tasks [24,39,1]. Asking respondents to complete a series of choice tasks, for example, is a fundamental element of the design of the increasingly popular set of SP elicitation formats known as discrete choice experiments (DCEs) [39,30].

The extent to which useful additional information is elicited by repeat-response formats is challenged by the growing body of evidence for *order effects* [16,33,21,8,29,19]; a term that embraces a variety of phenomena in which systematic changes in expressed preferences are observed along the sequence of valuation tasks. Order effects present SP practitioners with a profound problem; what do choices in a repeat-response SP exercise tell us about 'true' preferences when choice

* Corresponding author. Fax: +44 1603 591327.
E-mail address: Brett.Day@uea.ac.uk (B. Day).

0095-0696/$ - see front matter
doi:10.1016/j.jeem.2011.09.001

behavior appears to change when the valuation questions are simply presented in a different order? While order effects in simpler SP formats have been studied in detail [27,21,18,6], it is only comparatively recently that DCEs have come under the same scrutiny [13,29,36,19].

The present paper seeks to contribute to that growing literature in a number of ways. First, in Section 2 we attempt to classify the diverse array of order effect phenomena according to the pattern of preference change and the aspect of task order to which those changes relate. Moreover, we review a variety of theories regarding why order effects are observed in repeat-response SP data. We group those explanations according to the posited mechanism through which order effects arise and then use our classification of order effect phenomena to characterize the patterns suggestive of those different mechanisms.

Second, in Section 3 we consider how traditional DCE designs may fail to provide for the econometric identification of ‘actual’ order effects while potentially inducing the identification of ‘spurious’ order-dependence. We use those insights to inform the design of an empirical study taking the form of a DCE in which respondents make choices in a series of seventeen binary discrete choice tasks relating to the quality of domestic tap water.

One design factor that has been shown to be important in mitigating ordering effects in repeated contingent valuation (CV) questions, is whether the series of tasks are revealed to respondents before the exercise (so-called *advanced disclosure*) or whether they are revealed task by task as the exercise proceeds (so-called *stepwise disclosure*) [5]. The paper’s third contribution, therefore, is to investigate whether those two different presentational formats impact ordering effects in the context of a DCE.

In Section 4 we specify a preference function that admits each of the predicted patterns of order effect suggested by the various theories of behavior reviewed in Section 2. Separate models are estimated for the advanced disclosure sample and the stepwise disclosure sample. Examination of the parameters of those models allows us to identify the key patterns of order effect discernible in each samples’ responses and to assess the degree to which our data support the various competing theories of behavior. Comparison across the models allows us to assess how the two disclosure treatments impact on expressed preferences. In addition, our experiment includes a novel design element whereby respondents are presented with identical first and last choice tasks. As outlined in Section 5, comparison of responses to those repeated tasks allows for the identification of order effects using non-parametric statistics.

Our empirical study provides strong evidence for various forms of order effect in DCE data. We also find that the advanced disclosure of tasks results in significant differences in the expression of order effects. While those empirical findings are interesting in their own right, we hope that this paper’s additional contributions are in providing a systematic characterization of order effects in DCEs, and in proposing ways in which DCEs can be designed and analyzed so as to more thoroughly study those phenomena.

2. Ordering effects in DCE exercises

2.1. A definition and categorization of order effects

DCEs are a SP method used to elicit preferences for non-market goods. In a DCE respondents are asked to complete one or more *choice tasks*. Each task is constructed from two or more *options* where each option describes a particular manifestation of the non-market good in terms of the levels of its *attributes*. Typically those attributes are various quality dimensions of the non-market good and a cost of provision. Respondents are asked to indicate which of the options in each task is their most preferred. Respondents’ choices not only inform on the structure of their preferences for; (1) individual attributes and (2) particular options (especially, perhaps, options identified with a particular state of the world such as the status quo (SQ) level of provision of the good), but also on (3) the degree of variability in their choice behavior.

The standard economic model of demand for quality-differentiated goods assumes that in possession of complete information individuals form well-defined and stable preferences [37]. If respondents make choices in a DCE that truthfully manifest preferences of that type, then along a series of tasks their choices should be consistent with stable preferences for attributes and options and a constant degree of variability in choice behavior. Order effects are the observation of some systematic change in any of those three elements either relating to position in the sequence of tasks, henceforth *position-dependent order effects*, or relating to the nature of options in previous tasks, henceforth *precedent-dependent order effects*. With respect to the latter, the precedent may be set by previous levels of attributes or by some notion of the quality of the ‘deal’ offered in previous options. Moreover, ‘previous’ may refer to a particular task, for example to features of the first task observed, or it may refer to some sense of averaging over a range of previous tasks, for example the average cost observed thus far. Alternatively, the relevant comparator might be extreme features, for example the option offering the ‘best’ deal observed to date. Clearly, the term ‘order effect’ describes a diverse array of possible phenomena.

2.2. Causes of order effects

Evidence for various forms of order effect have been reported in a variety of repeated-choice contexts; in choices in SP exercises (e.g. [16,29,19]), in incentivized choices in experimental settings (e.g. [23,12]) and in real purchasing behavior (e.g. [32,45,40]). That literature proposes numerous possible causes for order effects, including strategic misrepresentation, referencing, institutional learning, changing preferences and changes in levels of cognitive effort. In the following, we briefly review those explanations and, as summarized in Table 1, describe the particular patterns of order effect characterizing each.

Table 1
Different theories of preference formation and the patterns of response they may exhibit in a repeat-response SP exercise.

Theory	Position dependence			Precedent dependence		
	Preferences for options	Preferences for attributes	Respondent randomness	Preferences for options	Preferences for attributes	Respondent randomness
Standard choice behaviors:						
'Standard' model	None	None	None	None	None	None
Confounded standard choice behaviors:						
Strategic misrepresentation	None	None	None	Judged by comparison to the best deals offered in previous options[a]	Judged by comparison to the best levels of attributes in previous options[a]	None
Institutional Learning	Converge after initial change[a]	Converge after initial change[a]	Decreasing randomness[a]	None	None	None
Failing credibility	Increasingly favor SQ[a]	None	Increasing randomness[a]	None	None	None
Fatigue effects	Increasingly favor SQ	Increasingly favor one attribute	Increasing randomness	None	None	None
Uncertain cost/income	Decreasing WTP for non-market commodity[a]		None	None	None	None
Non-standard choice behaviors						
Preference learning	None	Changes in preferences for non-cost attributes	Decreasing randomness	None	None	None
Anchoring effects	None	None	None	Judged by comparison to options in first task[a]	Judged by comparison to attribute levels in first task[a]	None
Reference effects	None	None	None	Judged by comparison to 'reference' deal developed from observation of deals in previous options[a]	Judged by comparison to 'reference' attribute level developed from observation of previous attribute levels[a]	None

[a] Order effects likely to be less prominent in advanced disclosure formats.

2.2.1. Causes of order effects: confounded 'standard' preferences

A *prima facie* interpretation of respondent choices in a DCE requires that those respondents have accepted a series of assertions made in the survey instructions. Typically respondents are asked to accept that each option represents the potential outcome of a real program of work. They are asked to treat each task as if it presented the only choice to be made. They are informed that their choices will help determine which particular program becomes reality. Of course, respondents may deliberately or inadvertently disregard such instructions. In that case, even if their preferences conform to the standard model, their choices may not.

For example, Carson and Groves [17] point out that presenting a series of options each offering the non-market commodity in some different configuration of price and qualities may undermine the credibility of those options as potential programs of work. Rather, individuals may interpret the exercise as one designed to guide decision-makers as to the optimal quality of provision and/or pricing. A rational economic agent interpreting the exercise in that way, should consider how best to answer the questions so as to manipulate the outcome to their advantage. Indeed, when the commodity is a pure public good, *strategic misrepresentation* of preferences may prove a dominant strategy. For example, to try to encourage a low cost of provision, respondents might reject options being offered at a cost that is higher than the lowest cost observed thus far in the exercise. Alternatively, as respondents progress through the exercise they may form a sense of what constitutes a good deal and reject options offering relatively poorer deals even when the truthful answer to the task would be to select that option. In essence, therefore, strategic misrepresentation would manifest itself as a precedent-dependent order effect in which the attributes/options offered in the current task are likely to be judged relative to the best attributes/options of previously observed tasks.[2]

Institutional learning represents another theory in which the elicitation procedure confounds the expression of standard preferences [44,12]. Here, initially confused respondents make choices that more reliably imitate their preferences as they work through the exercise gaining familiarity with the task format [51]. Institutional learning could account for position-dependent order effects in attributes, options or response variability, particularly those exhibiting a pattern of rapid initial change followed by stability.[3]

Carson and Groves [17] highlight the issue of informational *credibility* in SP exercises. In that regard, it is possible that successive presentation of assorted variants of the non-market commodity undermines the credibility of options as outcomes of real programs of work. A progressive loss of faith in the sincerity of the SP exercise could manifest itself in several ways such as an increasing tendency to reject options offering costly (and increasingly less credible) programs of improvements in favor of sticking with the SQ. Individuals might also simply stop making considered choices, a behavior that would likely manifest itself as increasing randomness in responses.

In their wide-ranging review, Carson and Groves [17] also note that the presentation of a second task with differently costed options may alert respondents to the prospect that the price they may be compelled to pay is uncertain. Provided this *uncertainty* is perceived as increasing the potential variance of future income, then from the second task onwards risk-averse respondents may exhibit declining willingness to pay (WTP) for the non-market commodity.

Finally, growing *fatigue* resulting from the cognitive burden of processing repeated choice tasks may undermine respondents' motivation to carefully consider their preferences in responding to each choice task. Fatigued respondents may, for example, begin to exhibit greater randomness in their choices (e.g. [11,14,48,49]). Alternatively, they may reduce the cognitive load by increasingly basing their decisions on only a subset of the attributes [52]. For example, always choosing the option with the lowest cost or highest level of a particular attribute [23,26]. Further, if the advantages of one option are relatively easy to assess (for example, the SQ option), then respondents may increasingly select that "safe choice" forgoing the effort of properly considering the deals offered by other options in the choice task.

2.2.2. Causes of order effects: non-standard preferences

A second set of theories postulate that order effects arise because individuals' preferences do not conform to the assumptions of the standard economic model. For example, rather than assuming prior, well-formed preferences, Plott [44] contends that preferences are learned through the process of repeated choice-making. DeSarbo et al. [20] and Bateman et al. [4] argue that *preference learning* may explain order effects in repeat-response SP data. In particular, respondents may alter and refine their preferences for quality dimensions of the non-market good as they take up the challenge of trading-off those attributes while working through a DCE exercise. In contrast, since individuals are well-versed in trade-offs involving money, preferences for the cost attribute are unlikely to exhibit changes resulting from learning.

Along similar lines, it has been postulated that individuals may come to an SP exercise with ill-formed and malleable preferences [47,53]. As a result, features of options in the opening task provide cues around which a respondent's sense of their own preferences coalesce such that options in subsequent tasks are regarded favorably or unfavorably according to

[2] Of course, other more complex strategies may be adopted, though we do not test for those here. As Carson and Groves [17] note, a strategic agent's optimal choice depends not only on their own preferences but also on their expectations of other respondents' preferences and on how the choices made by respondents are aggregated by the regulator to inform the decision.

[3] Clearly identifying a unique cause for different effects may be difficult. For example, institutional learning may lead respondents to realize that more than one configuration of the public good is under consideration, which in turn may motivate strategic misrepresentation.

how they compare to those in the initial task [2,3]. A substantial literature documents starting point effects in repeat-response SP exercises (e.g. [10,27,36]) that are consistent with this *anchoring hypothesis*.

An alternative account of order effects has its roots in Wicksteed's [54] proposition that buyers' behavior is significantly affected by whether they regard offers as being 'good deals' or 'bad deals'. In the marketing literature, extensive empirical evidence points to a process whereby consumers draw on their previous market experience to form a reference price [32,45,40]. Subsequent offers are framed as good or bad by comparison to that price. In a DCE such behavior would lead respondents to frame choices in a particular task by comparison to a reference developed from the features of previous tasks, behavior we describe as a *reference effect*. Extending the intuition of the marketing literature, respondents may form references based upon the best or worst attribute levels seen in previous tasks or as some expectation derived from that previous experience. Alternatively they may focus on features of tasks to which they attribute special informational importance. According to the social psychology literature, individuals may apply conversational or social norms in attributing such significance; attaching greatest importance to information presented first (primacy) or last (recency) in a sequence (e.g. [31,28]). In the case of primacy, features of the current task would be framed as 'good' or 'bad' by comparison with features of the first task. In the case of recency, that framing would be with regards to features of the previous task.

2.3. Order effects and task awareness

Conventionally, repeat-response SP exercises have adopted *a stepwise disclosure* (STP) format, in which respondents complete one SP task prior to being informed about a subsequent task which, once completed, is succeeded by a further unannounced task and so on. Alternatively, the full set of tasks can be revealed to respondents before they commence the exercise, a format described as *advanced disclosure* (ADV).

There are intuitive reasons to suspect that ADV formats may mitigate order effects. For example, since respondents know in advance that they are to face a series of tasks with varying attribute levels, issues associated with the progressive revelation of tasks (e.g. diminishing task credibility or increasing income uncertainty) are much less likely to be a factor in ADV formats. Likewise, respondents familiar from the outset with the scope of the exercise, may be less likely to form reference points relating to the features of particular tasks or attribute levels (e.g. first observed, best observed so far, etc.). Also, with STP formats information on the range of possible attribute levels is revealed to respondents only through the observation of successive tasks. As a result, a strategizing respondent's perception of what may be an optimal choice will depend on the order of task presentation. Such considerations should not be a factor in the ADV format where the entire strategy space is known to the respondent from the outset. In Table 1, patterns of order effect predicted by particular theories that we speculate are likely to be less prominent in ADV formats are identified with an asterisk.

Of course the possible mitigation of order effects in ADV format exercises does not necessarily translate to choices being more revealing of 'true' preferences. Rather, the reactions to progressive task revelation that lead to position dependent order effects in STP formats could, in an ADV format, simply be played out in their entirety before the exercise commences. Likewise, respondents to an ADV format may be just as liable to precedence dependence in their responses, but the cues that shape their responses are drawn from the initial revelation of the full set of tasks and, as a result, could remain constant over the exercise.

Relatively, few studies have addressed the impact of ADV formats on order effects. Bateman et al. [5] describe a CV study where respondents were asked to value three nested levels of environmental improvement. Using the STP format, the values attributed to those three goods differed significantly according to the order of their presentation. Those differences were not apparent using the ADV format, suggesting a mitigation of *position-dependent order effects*. In contrast, Bateman et al. [6] found that an ADV format failed to alleviate *precedent-dependent order effects* in a DCE exercise involving two binary choice tasks. In particular, second task responses were observed to differ systematically according to the deal offered in the initial task; a pattern redolent of the *starting task effects* described previously.

3. Experimental design

3.1. Isolating order-effects

In order to focus on the issue of order effects, the empirical experiment used in this research takes the form of a deliberately simple DCE. For example, to avoid confounding factors relating to unfamiliarity with complex, multi-attribute goods [38,22,7], our study concerns the everyday commodity of domestic tap water. It also focuses on just two quality attributes of that commodity: discoloration (due to soil run-off) and odor/taste problems (due to chlorination). Surveys indicate that consumers in the study area (Norfolk, UK) have experience with variation in these attributes [42]. Moreover, since they are also used to paying for service changes through their water bills, improvements in those quality aspects can be associated with a cost through a familiar and coercive payment vehicle. The attributes are described in terms of: (1) days annually where a household's tap water smells and tastes of chlorine (SMELL), (2) days annually where tap water is a rusty color (COLOR) and (3) the addition to the household's annual water bill induced by implementing technical fixes for these quality problems (COST).

	'No Scheme'	'Scheme A'
Number of days each year on which your tap water smelled and tasted of chlorine	10	3
Number of days each year on which your tap water was a rusty color	5	1
Addition to your annual water bill	£0	£10
	Choose 'No Scheme'	Choose 'Scheme A'
Which would you choose? (*tick one box only*)	☐	☐

Fig. 1. Typical choice question.

In addition, each task in our experiment (Fig. 1) is a simple binary choice between a constant SQ option in which no intervention is undertaken and costs remain unchanged, and a single 'alternative' state of the world, varied across tasks. Following discussion with water quality scientists[4] regarding the feasibility of preventative works, four levels were assigned to each attribute: SMELL (0, 3, 6 and 10 days); COLOR (0, 1, 3, and 5 days), and COST (£10, £20, £30 and £50). Since each alternative is only ever paired with the SQ option, our design also controls for preference anomalies that may result from within-task comparisons [9].

3.2. *Position-dependent order effects: parametric identification*

For the most part, the large literature advising on the optimal design of DCEs (e.g. [35,39,34]) is predicated on the assumption that preferences are unaffected by task order. Such designs will likely prove inadequate for the identification of order effects. In our study, the full factorial design contains $4^3=64$ unique options. Each of those options, when paired with the SQ, constitutes a choice task for our experiment. To introduce high-level order variation into the design, those 64 tasks were divided into two groups. The first group contained the 'extreme' options; that is, those $2^3=8$ options constructed from only the highest and lowest attribute levels. The second group comprised the remaining 56 'intermediate' options. Half the sample was presented with 'extreme' tasks followed by 'intermediate' tasks, while for the other half that order was reversed. To introduce order variation at the lower level, eight different order variations of the 'extreme' choice tasks were generated using a Latin square design [50]. The 56 'intermediate' tasks were randomly divided into eight sets, each containing seven choice tasks. Each order variation of the 'extreme' tasks was paired with a different set of 'intermediate' tasks giving 8 blocks consisting of 15 choice tasks. Reversing the order of presentation of 'extreme' and 'intermediate' tasks gave a final design with 16 blocks.

Identification of position-dependent order effects depends not only on suitable variability in the order of task presentation, but also on maintaining balance in the design at all positions in the task sequence. In a parametric analysis, failure to maintain such balance may lead to design-induced position dependence in the scale of the error term, a pattern that might incorrectly be attributed to changes in choice behavior. To elucidate, in the random utility model the error term captures both respondent randomness and residuals resulting from misspecification of the preference function. Unless the model correctly imitates the true data generating process, the analyst's specification will do relatively better at predicting the preference function over certain ranges of the attribute space and relatively worse elsewhere. Accordingly, unless the design maintains a constant coverage of the attribute space at each position, the misspecification component of the error term might differ in scale along the sequence of tasks confounding identification of position dependence in respondent randomness. In our study, we attempt to minimize such confounding in two ways. First, our full-factorial design allows us to estimate interaction effects in preferences for attributes, avoiding a known source of misspecification bias [43]. Second, the variation in task order in our 16 blocks was chosen to ensure roughly the same mix of attribute levels across the sample at each position in the sequence.[5]

[4] We are grateful to Professor Glen George, Centre for Ecology and Hydrology, Lancaster, UK.

[5] While greatly improving on existing studies, our implementation still failed to ensure equivalent covariance matrices across all sequence orders. The substantive problem occurs where the 7 tasks from the 'intermediate' design were joined to the 8 tasks from the 'extreme' design. Recall that half the sample saw intermediate tasks followed by extreme tasks and half the sample the reverse. Unfortunately, the imbalance in the number of tasks in those two designs meant that at sequence position 8 only extreme tasks were presented. As a result, middle levels of all attributes are missing from options in that position.

3.3. Position-dependent order effects: non-parametric identification

We defined two levels of tap water quality; one a small improvement (SMALL) over the SQ, the other a relatively large improvement (LARGE).[6] Eight different tasks were constructed where the alternative option was one of the two improvements offered at one of the four levels of COST. Each respondent was randomly assigned one of those eight tasks, which was added to the sequence of tasks in both the first and last positions. Direct comparison of respondents' choices in this repeated task provides for a within-person non-parametric test of order effects.

Adding the repeated task has implications for the parametric identification of position-dependent order effects. Since those tasks exhibit little variation in the COLOR and SMELL attributes it is not possible to examine how tastes for those attributes differ in the first and last positions. In addition, while coverage of the attribute space is exactly equal in the first and last tasks and approximately equal for each of the intermediary 15 tasks, it differs substantially between those two groups of positions. As such, while differences in the error scale within those two task sets can be safely interpreted as position-dependent changes in behavior, comparison across them is confounded by possible differences in the extent of misspecification bias.

3.4. Precedent-dependent order effects: parametric identification

Several of the theories reviewed in Section 2 suggest that features of previously observed tasks may influence individuals' choices in the current task. The pertinent features may be the levels of individual *attributes* or the overall quality of individual *options*. In this application, we focus on the identification of precedence-dependence in options. To investigate that possibility we require a definition of the quality of 'deal' offered by an option; a definition provided in Section 4.4. Our experiment is designed such that, across the sample, 'extreme' options offering particularly good and bad deals are located at all positions from tasks 2 to 16. Likewise, eight different options are presented as the initial task and there are 16 different task orders across the middle 15 tasks. As a result, our data provide a wealth of variation from which to identify how the best or worst prior deal, or the first or preceding deal impact on current choices.

3.5. Order effects and task awareness

Our design also addresses the issue of choice task awareness [5]. Respondents were randomly allocated into either a STP or an ADV treatment group. The two treatments differed in that prior to the initial task the ADV group were shown the full range of attribute levels and each of the tasks they would subsequently face in the DCE. The impact of advanced disclosure can be investigated through comparison of choices made by those two treatment groups.

3.6. Survey administration

The survey was administered as face-to-face interviews at respondents' homes and took on average 21 min to complete (std. dev. 6.42 min). Addresses were selected at random from the study area and 53% of selected households agreed to participate, providing a sample of 864 respondents. Six questionnaires were subsequently dropped due to missing data.

4. Parametric econometric model

4.1. The random utility model with standard preferences

Our DCE presents respondents ($i=1, 2, \ldots, N$) with a series of binary choice tasks ($j=1, 2, \ldots, J$) pitting the *status quo* (SQ) against a costly *alternative*. The standard econometric framework for the analysis of such data is provided by McFadden's random utility model [41]. The usual point of departure is to assume that the utility derived by respondent i from option $\oplus$ in task j can be approximated by a linear indirect utility function of the form

$$U_{\oplus,ij} = \alpha_{\oplus} + \boldsymbol{X}'_{\oplus}\boldsymbol{\beta} + e_{\oplus,ij} \tag{1}$$

where $\alpha_{\oplus}$ is the fixed utility derived from choosing option $\oplus$, $\boldsymbol{X}_{\oplus}$ is the M-vector of option $\oplus$'s attribute levels and their cross-products, $\boldsymbol{\beta}$ is the vector of associated taste parameters, and $e_{\oplus,ij}$ is a residual element ensuring an identify between respondent i's actual evaluation of the utility of option $\oplus$ ($U_{\oplus,ij}$) and the model's approximation ($\alpha_{\oplus} + \boldsymbol{X}'_{\oplus}\boldsymbol{\beta}$).

If the SQ option is identified by subscript 0, then in choice task j respondent i is assumed to prefer the alternative offered by option $\oplus$ if $U_{\oplus,ij} > U_{0,ij}$ or, from (1), when

$$(\alpha_{\oplus} - \alpha_0) + \boldsymbol{x}'_{\oplus}\boldsymbol{\beta} > \varepsilon_{ij} \tag{2}$$

where $\boldsymbol{x}_{\oplus} = \boldsymbol{X}_{\oplus} - \boldsymbol{X}_0$ and $\varepsilon_{ij} = e_{\oplus,ij} - e_{0,ij}$. In this application, the ε_{ij} are taken to be independent draws from a mean-zero normal distribution with variance σ^2.

[6] The LARGE commodity offered an improvement of 7 less days of ODOR and 2 less days of COLOR. The SMALL commodity offered 4 less days of ODOR but no COLOR improvement.

4.2. Accounting for individual heterogeneity

Observe from (2) that the attribute taste parameters, $\boldsymbol{\beta}$, are assumed to be constant across individuals. We maintain that simplification in our application such that our analysis can be thought of as identifying tastes for the attributes at the mean of the data. In contrast, by assuming α_0 to be normally distributed in the population, we allow for the possibility that individuals may differ in their desire to switch away from the SQ, what we shall term an *SQ effect*. In practice that means replacing $-\alpha_0$ with $\alpha+\alpha_i$ in (2), where α is the mean of the distribution of the SQ effect and α_i is a centered normal variate with variance s_α^2. We also allow for the possibility that the scale of the error distribution may vary across individuals, perhaps reflecting differences in concentration and the precision with which respondents' choices reflect their preferences. Accordingly, individual error scales, σ_i, are assumed to be drawn from a log-normal distribution with location σ and scale s_σ^2.

To fix the measure of utility we make two further assumptions. First, given their generic nature, we assume that all alternative options have identical fixed utility and establish the location of the utility measure by normalizing that quantity to zero, such that $\alpha_{\mathbb{O}}=0$ $(\forall \mathbb{O}, \mathbb{O}\neq 0)$. Second, we fix the scale of the utility measure through the normalization $s_\alpha=1$. With those changes Eq. (2) can be rewritten as

$$\alpha+\alpha_i+\boldsymbol{x}'_{\mathbb{O}}\boldsymbol{\beta} > \varepsilon_{ij}(\sigma_i) \tag{3}$$

4.3. Position-dependent ordering effects

In assuming that preferences for attributes and options are constant across tasks, (3) is consistent with the assumptions of the standard model. Of course, alternative theories predict that taste parameters, $\boldsymbol{\beta}$, the SQ effect, α, and the scale of the error distribution, σ, may vary along the sequence of tasks (see Table 1). To that end, we amend (3) to allow those parameters to develop in a piecewise-constant fashion as the tasks progress. The 15 tasks forming the core of the design are divided into three, 5-task pieces and a dummy-coding specification used to allow for preference changes from piece to piece. Two further dummy variables are included to accommodate the repeated task appended to the start and end of the design. Hence, the mean of the SQ effect distribution is parameterized as $\alpha_1=\alpha+a_1$ in the first task, as $\alpha_j=\alpha+a_2$ $(j=2,\ldots,6)$ in the next five tasks, as $\alpha_j=\alpha+a_3$ $(j=7,\ldots,11)$ over the middle five tasks, as $\alpha_j=\alpha+a_4$ $(j=12,\ldots,16)$ in the eleventh to sixteenth tasks and as $\alpha_{17}=\alpha+a_5$ in the final task. Position-dependence in the location of the ecale dtribution is specified identically; $\sigma_1=\sigma+t_1$, $\sigma_j=\sigma+t_2$ $(j=2,\ldots,6)$, $\sigma_j=\sigma+t_3$ $(j=7,\ldots,11)$, $\sigma_j=\sigma+t_4$ $(j=12,\ldots,16)$ and $\sigma_{17}=\sigma+t_5$. For identification, the preferences expressed in tasks 2 to 6 are established as the baseline through the normalizations $a_2=0$ and $t_2=0$.

Since only two water quality improvements are used in the repeated task, our experiment does not allow for identification of a separate first and last task effect for the taste parameters. Accordingly, we adopt a restricted specification for each taste parameter, $m=1,\ldots,M$, in which position dependence is allowed through the parameterization; $\beta_{m,j}=\beta_m+b_{m,1}$ $(j=1,\ldots,6)$ in the first six tasks, $\beta_{m,j}=\beta_m+b_{m,2}$ $(j=7,\ldots,11)$ over the middle five tasks and $\beta_{m,j}=\beta_m+b_{m,3}$ $(j=12,\ldots,17)$ over the final six tasks. Here, the first 6 tasks are established as the baseline through the normalization $b_{m,1}=0$ $(m=1,\ldots,M)$. In the pursuit of parsimony, $b_{m,k}$ coefficients for the cross-product terms are assumed to be zero; our specification only allows for position-dependence in the attribute main effects.

4.4. Precedent-dependent ordering effects

In comparing options across different tasks we assume that respondents form a sense of the 'deal' being offered in each task. That deal is defined as the quantity of tap water quality improvements per £ of cost. Since respondents may regard SMELL and COLOR improvements differently, our measure of a 'deal' uses the taste parameters to form a preference-weighted sum of days of improvements. In particular, a respondent appraising the options in task j, is assumed to evaluate the deal offered in some previous task, k, as

$$p_{k,j} = -\frac{x_{SML,k}+\frac{\beta_{COL,j}}{\beta_{SML,j}}x_{COL,k}}{x_{CST,k}} \quad j=1,\ldots,J; \quad k=1,\ldots,j \tag{4}$$

Observe from (4) that an option offering no tap water improvements has a deal value of zero, while the negative sign preceding the right hand side expression ensures that the deal variable is increasing in days of tap water improvements but decreasing in costs.

Then for each task we construct the following vector of precedent-dependent variables:

$$\boldsymbol{P}_j = (\,p_{j,j}-p_{1,j} \quad p_{j,j}-p_{j-1,j} \quad p_{j,j}-p_j^- \quad p_{j,j}-p_j^+\,) \quad j=1,2,\ldots,J \tag{5}$$

where the first element of $\boldsymbol{P}_j$, records the difference between the deal in the current task and the deal in the first task allowing for *starting task effects*. The second element records the difference between the deal in the current task and that offer by the immediately preceding task allowing for *previous task effects*. The third and forth elements allow for the possibility that individuals may evaluate options with *reference to the extreme deals*; that is to say, the worst (p_j^-) and best (p_j^+) deals observed thus far in the series of tasks.

The precedent-dependent variables are associated with a vector of coefficients, $\boldsymbol{\gamma}=(\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4)'$, and included in our specification of the utility function in the additive form $\boldsymbol{P}_j\boldsymbol{\gamma}$. A useful way to interpret those coefficients is revealed by rearranging $\boldsymbol{P}_j\gamma$ to give

$$gp_{j,j}-(\gamma_1 p_{1,j}+\gamma_2 p_{j-1,j}+\gamma_3 p_j^-+\gamma_4 p_j^+) \quad j=1,2,\ldots,J \tag{6}$$

where $g=\gamma_1+\gamma_2+\gamma_3+\gamma_4$. Our specification can be seen to describe the utility impact of the levels of preceding deals as a linear function of the deal in the current task, $p_{j,j}$, with slope g. Setting (6) equal to zero allows one to solve for the 'reference deal'

$$p_j^*=\frac{\gamma_1}{g}p_{1,j}+\frac{\gamma_2}{g}p_{j-1,j}+\frac{\gamma_3}{g}p_j^-+\frac{\gamma_4}{g}p_j^+ \quad j=1,2,\ldots,J \tag{7}$$

This reference deal can be thought of as the deal to which a respondent makes comparison when evaluating the deal offered by the current task. If, as might be expected, $g>0$, then current deals that are increasingly better (worse) than the reference deal are viewed in a progressively more (less) favorable light. Observe that the reference deal, p_j^*, is simply a weighted average of the precedent deals, with the terms γ_h/g $(h=1,2,3,4)$ defining the weight put on each particular precedent. In other words, our specification allows us to identify how respondents combine the information provided by the first, previous, best and worst deals in forming their conception of the reference deal.

Our final specification of the choice rule is; choose option $\oplus$ in choice task j if;

$$\alpha_j+\alpha_i+\boldsymbol{x}'_{\oplus}\boldsymbol{\beta_j}+\boldsymbol{\gamma}\boldsymbol{P}_j>\varepsilon_{ij}(\sigma_{ij}) \tag{8}$$

4.5. Estimation

The econometric model we employ in order to recover the parameters of the preference function in (8) is essentially a random effects probit model generalized through the assumption that the error scale is also treated as a random parameter with a log-normal distribution. The likelihood of observing the set of J choices made by respondent i is (see [25])

$$l_i=\int_{-\infty}^{\infty}\phi(\alpha_i)\prod_{j=1}^{J}\int_0^{\infty}\phi\left(\frac{\ln\sigma_i-\sigma_j}{s_\sigma}\right)\Phi\left(q_{ij}\left(\frac{v_j}{\sigma_i}-\sqrt{\frac{\rho_j}{1-\rho_j}}\right)\right)d\sigma_i\,d\alpha_i \tag{9}$$

where, y_{ij}, is the dependent variable that records a value 0 if inidividual i chose the SQ option in task j and 1 if the alternative option was chosen, $q_{ij}=2y_{ij}-1$, $v_j=\alpha_j+\boldsymbol{x}'_j\boldsymbol{\beta_j}+\boldsymbol{\gamma}\boldsymbol{P}_j,\rho_j=1/(\sigma_i^2+1)$ and $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative and density function of a standard normal distribution respectively. The outer integral over α_i. can be calculated using Gaussian quadrature [15] while simulation methods are used to approximate the inner integral over σ_i. The simulated log-likelihood function to be maximized with respect to the parameters is then

$$\ln L(\alpha,\boldsymbol{a},\sigma,s_\sigma,\boldsymbol{t},\boldsymbol{\beta},\boldsymbol{b},\gamma)=\sum_{i=1}^{N}\ln l_i \tag{10}$$

5. Results: nonparametrics

5.1. Stated demand for the small and large improvements

Before presenting the results of the parametric analysis, let us first consider the direct test of order effects provided by comparison of choices in the repeated first and last task. Recall that eight different alternative options were used in the repeated task; a SMALL or LARGE water quality improvement offered at one of four possible costs (£10, £20, £30, or £50). Respondents were randomly assigned to a subsample receiving one of those eight options in their repeated task.[7] A measure of *stated demand* in each subsample is provided by the proportion of respondents in that subsample choosing the alternative option. In Fig. 2 those proportions are plotted on two graphs, one for First Task responses and one for Last Task responses. To aid interpretation, on each graph the point observations of stated demand for the SMALL and LARGE improvements are joined by line segments that approximate the path of the willingness to pay survivor (WTP) functions for those two commodities.

Notice from Fig. 2 that the line segments of the WTP survivor functions are generally downward sloping; stated demand tends to decline monotonically with cost. The only exception is the upward turn of the WTP survivor function for the SMALL good offered at £50. For that case, as for all other cases, however, a two-tailed z-test of the equality of proportions in independent samples confirms that there are no statistical violations of monotonicity.

The data illustrated in Fig. 2 also indicate that, all else equal, individuals prefer a large improvement to a small one. We compare the set of point estimates of the WTP survivor function for the SMALL good to those of the LARGE good and confirm statistically that stated demand for the latter is significantly greater (first task: χ^2-stat 26.29, p-value <0.001; last

[7] In an analysis available on request, the socioeconomic characteristics of the sub-samples were compared and no statistically significant differences found.

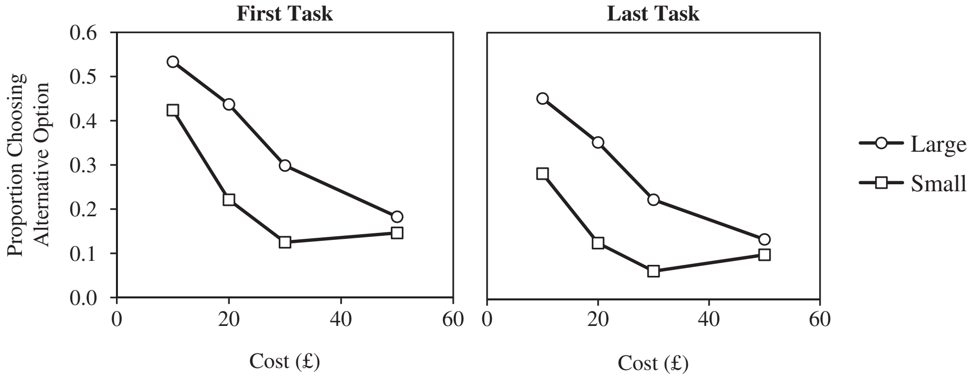

Fig. 2. Stated demand for the small and large improvements for the whole sample.

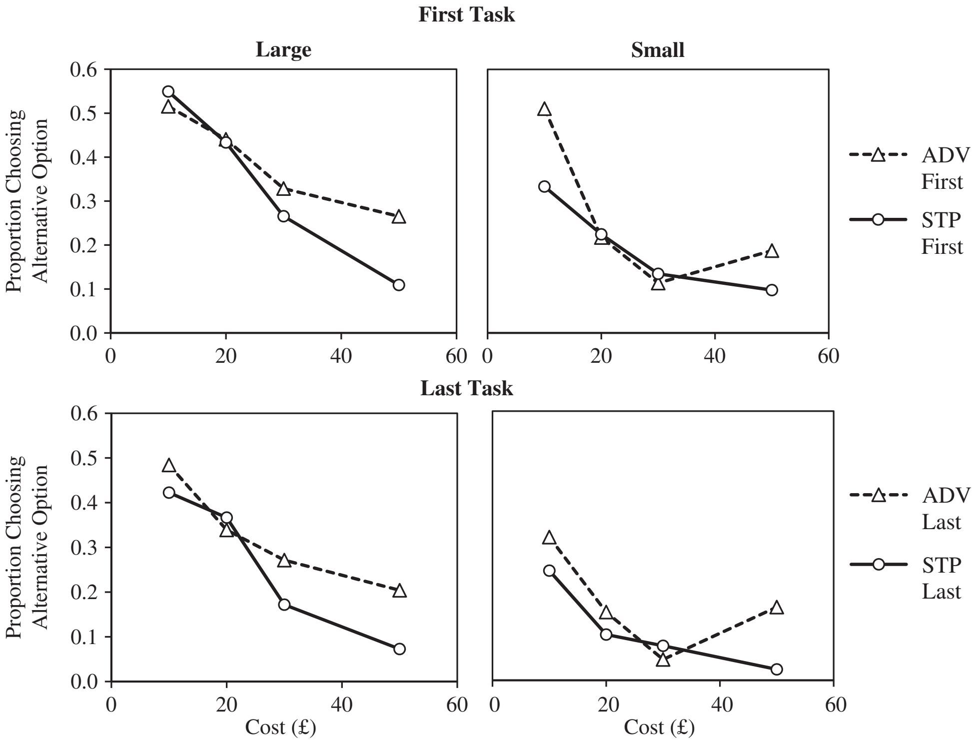

Fig. 3. Stated demand for the small and large improvements in the STP and ADV treatment groups.

task: χ^2-stat 37.79, p-value <0.001). All in all, Fig. 2 relates a satisfying story. Whether individuals are answering the very first task or the seventeenth and final task, their responses closely conform to some fundamental predictions of economic rationality.

Recall that respondents were also randomly assigned to an advanced (ADV) or stepwise (STP) revelation treatment group. Fig. 3 repeats the data in Fig. 2, but plotting the two treatment groups separately. Notice that in all cases the point estimates of the WTP survivor functions for the two treatment groups are either virtually identical or stated demand for the ADV group is noticeably higher. In two cases those differences are found to be significant at the 90% confidence level (LARGE last at £50: z-stat 1.9591, p-value 0.0501; SMALL first at £10: z-stat 1.7203, p-value 0.0854) and in two further cases significant at the 95% confidence level (LARGE first at £50: z-stat 2.0579, p-value 0.0396; SMALL last at £50: z-stat 2.1871, p-value 0.0287). Accordingly, there is some evidence to support the idea that WTP for tap water quality improvements is higher in the ADV treatment group than in the STP treatment group.

5.2. Stated demand in the first and last task

Consider Fig. 4 which plots the same data as Fig. 2 but in such a way that responses to the first and last tasks can be directly compared. It is immediately evident that in all cases there is a marked downward shift in the proportion of

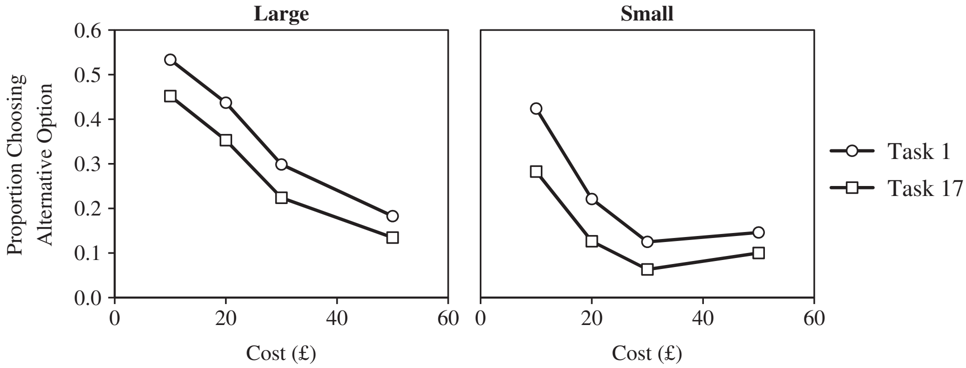

Fig. 4. Stated demand for the small and large improvements in the first and last task for the whole sample.

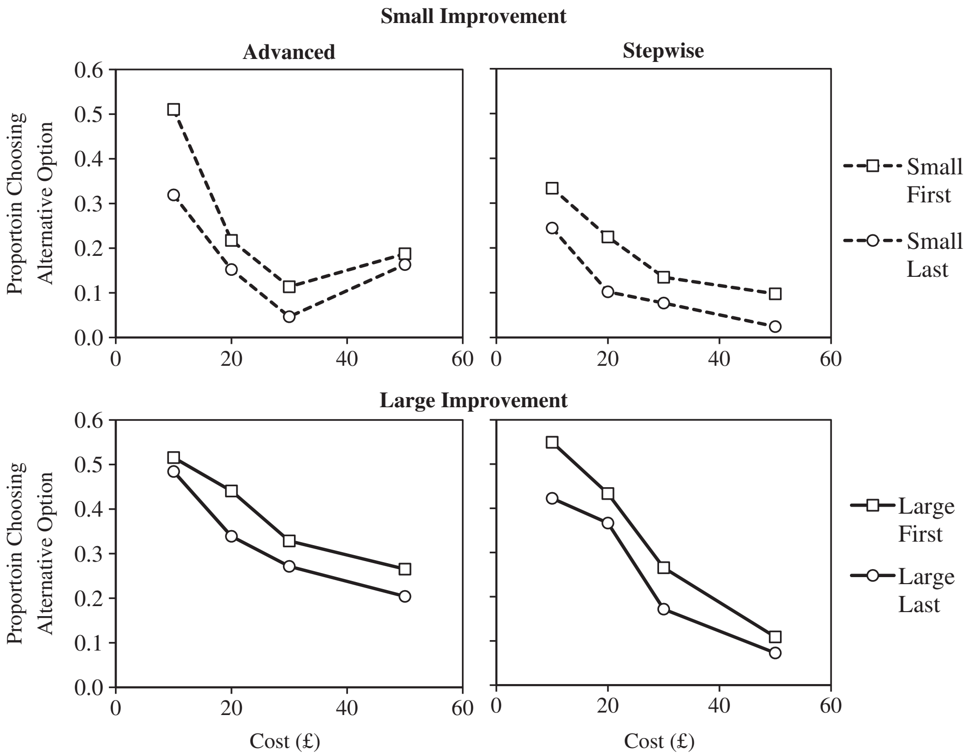

Fig. 5. Stated demand for the small and large improvements in the first and last task in the STP and ADV treatment groups.

respondents choosing the alternative option in the last task. Note that the comparisons in these graphs are within-person; that is to say, they compare the choices made by one subsample in the first task to the choices made by that same group when evaluating the identical task in the final question. The appropriate statistical test is McNemar's test of change which confirms that the observed shift in preferences is highly significant both for the SMALL (χ^2-stat 21.84, p-value <0.001) and LARGE (χ^2-stat 16.75, p-value <0.001) improvements. Our data provides unequivocal evidence of order effects in expressions of preference.

Fig. 5 presents a similar analysis though the data is now further partitioned into ADV and STP treatments. The conclusions of the previous analysis still hold; in both STP and ADV groups there is a systematic and substantive downward shift in stated demand between the first and last task. In all cases, McNemar's test reveals that shift in expressed preferences to be statistically significant (STP SMALL: χ^2-stat 11.25, p-value <0.001; STP LARGE: χ^2-stat 11.28, p-value <0.001; ADV SMALL: χ^2-stat 9.38, p-value 0.002; ADV LARGE: χ^2-stat 5.30, p-value 0.021). It appears that advanced revelation of choice tasks does not eliminate order effects in expressions of preference.

Fig. 5 also draws attention to the fact that the upward turn of the WTP survivor function for the SMALL good offered at £50 is isolated to the ADV treatment. It is interesting to note that this apparent anomaly is expressed by the subsample aware of the fact that they were being asked to pay the highest price for the SMALL good; a combination that may have lead them to question the credibility of this option.[8]

6. Results: parametrics

While our non-parametric analysis provides unambiguous evidence of a significant difference between the preferences expressed in the first and last tasks, it does not reveal whether or how preferences change over the sequence of intervening tasks. Nor does it clarify the mechanisms driving such changes. Our parametric analysis seeks to address those questions.

First, note that 19.3% of the STP sample and 16.5% of the ADV sample selected the SQ option in all 17 tasks. The choices of these corner solution respondents reveal little regarding how preferences change along the sequence of choice tasks. This pattern of choices is also redolent of so-called protest behavior; that is to say, it could signal a refusal on the part of the respondent to make choices that involve trading-off between attributes. Accordingly, respondents with this pattern of choice behavior are not included our parametric models.[9]

Table 2 presents coefficient estimates from two models; one estimated from the responses of the STP treatment group and the other from those of the ADV treatment group. The estimates are expressed in units of utility though the arbitrary normalizations used to fix the scale of utility ($\alpha_{\oplus} = 0 \; \forall \; \oplus \neq 0$, $\sigma_{\alpha}^2 = 1$) do not guarantee that those units are the same in both models. As a result, individual parameter estimates should not be directly compared across the two models. It is relatively easy, however, to compare the models as a whole. A likelihood ratio test reveals that we can reject the hypothesis that the two groups express identical preferences with greater than 99% confidence (*ln L*(constrained): −4206.9, *LR*-stat: 52.28, *p*-value: 0.0024).

6.1. Attribute main effects

The first set of parameters in Table 2, describe the main effects of the experimental attributes; COST, COLOR and SMELL. For both the ADV and STP models, each of those main effects (β_{CST}, β_{COL}, β_{SML}) is negative and statistically significant; as expected, extra cost and extra days of tap water problems are viewed as a bad thing. Notice that in both treatment groups a day of COLOR problems is regarded as roughly twice as bad as a day of SMELL problems.

Recall that position dependence in the attribute main effects is modeled using a piecewise-constant specification taking tasks 1–6 as the baseline and estimating parameters to capture shifts in preferences over tasks 7–11 and over tasks 12–17. In both models we find that the shift parameters for the COST coefficient are small in absolute value and lack statistical significance. It appears that respondents in both ADV and STP treatments groups maintain a constant assessment of the marginal disutility of expenditure over the sequence of choice tasks.

For the ADV treatment a similar conclusion can be reached with regards to preferences for the two water quality attributes. Not so for the STP treatment. As the STP group progresses through the sequence of tasks, the disutility attributed to days of both water quality problems is seen to increase markedly. That increase is most substantial over the middle 5 tasks; a shift in preferences that proves significant at 95% level of confidence for the SMELL attribute and the 99% level of confidence for the COLOR attribute.

To give a clearer impression of how the patterns of preferences expressed by the two treatment groups differ, Fig. 6 plots out implied marginal willingness to pay (MWTP) for tap water improvements. MWTP is calculated as the ratio of the main effect of a tap water attribute to the COST attribute and, since it is expressed in the money metric, circumvents difficulties of directly comparing parameters expressed in the different utility scales of the two models.

Observe that in the first six tasks the STP group exhibit relatively low MWTP for attributes. Tests using resampling methods [46] confirm those values to be significantly lower than those expressed by the ADV group (SMELL *p*-value: 0.0040; COLOR *p*-value: 0.0174). In the STP group, MWTP for both attributes increases sharply in the middle series of tasks. Indeed, for both the SMELL and COLOR attributes we cannot reject the hypothesis that the two groups express identical values over this middle range of tasks (SMELL *p*-value: 0.4890; COLOR, *p*-value: 0.3944). Over the final six tasks that comparability of values remains statistically supported for the SMELL attribute (*p*-value: 0.1915), but for the COLOR attribute MWTP for the STP group declines to a value that is lower than that of the ADV group by an amount that is statistically significant, but only at the 90% level of confidence (*p*-value: 0.0577).

Taken as a whole, our models suggest that under the ADV treatment MWTP for both tap water attributes remains stable, in a statistical sense, across the sequence of tasks. In contrast, MWTP values under the STP treatment start out comparatively low but rise sharply as the exercise progresses converging on the values of the ADV group. For the SMELL attribute that convergence is maintained through to the end of the exercise. For the COLOR attribute MWTP is somewhat lower in the STP group than the ADV group over the final 6 tasks.

---

[8] Unfortunately, sample sizes are too small at this level of disaggregation of the data to provide strong statistical support for this finding. The up-turn does not result in a statistical violation of monotonicity of the WTP survivor function in the first task (*z*-stat 0.8953, *p*-value 0.3245), though there is stronger evidence for a statistical violation in the last task (*z*-stat 1.7951, *p*-value 0.0726). In addition, the up-turn leads to stated demand from the ADV subsample valuing the SMALL good converging on that of the ADV subsample valuing the LARGE good (first task: *z*-stat 0.9149, *p*-value 0.3602; Last task: *z*-stat 0.9853, *p*-value 0.6018); we cannot reject the possibility that the ADV treatment group are insensitive to scope at the highest price point.

[9] Results from the full sample are qualitatively similar in all respects though, as might be expected, the magnitude of implied values for attributes are somewhat diminished through the inclusion of respondents that consistently chose the SQ. Results of the model estimated from the full sample are available from the lead author on request.

Table 2
Parametric models of choices of STP and ADV treatment groups allowing for position dependence and precedence dependence in preferences.

Variable	STP		ADV	
	Param (std. err.)	*p*-value	Param (std. err.)	*p*-value
Main effects				
COST coefficient (β_{CST})	−0.0334*** (0.005)	< 0.0001	−0.0332*** (0.005)	< 0.0001
Tasks 1–6 ($b_{CST,1}$)	0	–	0	–
Tasks 7–11 ($b_{CST,2}$)	−0.0023 (0.005)	0.6055	−0.0006 (0.004)	0.8726
Tasks 12–17 ($b_{CST,3}$)	0.0035 (0.004)	0.3332	0.0042 (0.003)	0.2000
SMELL coefficient (β_{SML})	−0.0678*** (0.018)	0.0001	−0.1062*** (0.018)	< 0.0001
Tasks 1–6 ($b_{SML,1}$)	0	–	0	–
Tasks 7–11 ($b_{SML,2}$)	−0.0410** (0.016)	0.0114	0.0019 (0.014)	0.8936
Tasks 12–17 ($b_{SML,3}$)	−0.0198 (0.014)	0.1568	0.0076 (0.013)	0.5494
COLOR coefficient (β_{COL})	−0.1252*** (0.032)	0.0001	−0.1796*** (0.031)	< 0.0001
Tasks 1–6 ($b_{COL,1}$)	0		0	–
Tasks 7–11 ($b_{COL,2}$)	−0.0925*** (0.029)	0.0014	−0.0365 (0.026)	0.1563
Tasks 12–17 ($b_{COL,3}$)	−0.0206 (0.023)	0.3713	−0.0122 (0.021)	0.5622
Interaction effects:				
Color × Smell ($\beta_{COL\times SML}$)	−0.0187*** (0.003)	< 0.0001	−0.0089*** (0.003)	0.0020
Color × Cost/10 ($\beta_{COL\times CST}$)	0.0099 (0.007)	0.1606	0.0144** (0.007)	0.0258
Smell × Cost/10 ($\beta_{SML\times CST}$)	−0.0007 (0.004)	0.8680	0.0013 (0.004)	0.7110
SQ effect: $\sim N(\alpha_j, s_\alpha)$				
Scale of distribution (s_α)	1		1	–
Location of distribution (α)	−1.4076***158)	< 0.0001	−1.0645*** (0.144)	< 0.0001
Task 1 (a_1)	0.5664*** (0.100)	< 0.0001	0.3710*** (0.090)	< 0.0001
Tasks 2–6 (a_2)	0		0	–
Tasks 7–11 (a_3)	−0.6746*** (0.179)	0.0002	0.0322 (0.159)	0.8396
Tasks 12 to 16 (a_4)	−0.5465*** (0.176)	0.0019	0.0798 (0.155)	0.6074
Task 17 (a_5)	−0.1322 (0.171)	0.4403	0.5788*** (0.160)	0.0003
Precedent dependent effects:				
Start deal (γ_1)	0.3724** (0.156)	0.0173	0.3994*** (0.149)	0.0075
Previous deal (γ_2)	−0.0595 (0.070)	0.3918	0.0372 (0.058)	0.5213
Best deal (γ_3)	0.1351** (0.067)	0.0448	0.3281*** (0.069)	< 0.0001
Worst deal (γ_4)	0.5427*** (0.159)	0.0006	−0.0474 (0.139)	0.7327
Error scale: $\sim LN(\sigma_j, s_\sigma)$				
Scale of distribution (s_σ)	0.5473*** (0.057)	< 0.0001	0.3403*** (0.049)	< 0.0001
Location of distribution (σ)	−0.2979*** (0.087)	0.0007	−0.2206*** (0.077)	0.0043
Task 1 (t_1)	−0.3589** (0.170)	0.0350	−0.2731* (0.151)	0.0698
Tasks 2–6 (t_2)	0	–	0	–
Tasks 7–11 (t_3)	0.0574 (103)	0.5777	0.0822 (0.092)	0.3737
Tasks 12–16 (t_4)	−0.0400 (100)	0.6885	−0.1699* (0.093)	0.0675
Task 17 (t_5)	−0.4034** 0.181)	0.0261	−0.5652*** (0.175)	0.0013
ln *L*	−1994.402		−2186.346	
Num respondents (*N*)	349		355	
Tasks per respondent	17		17	

* Significant at 90% confidence level.
** Significant at 95% confidence level.
*** Significant at 99% confidence level.

6.2. Attribute interaction effects

In both models we observe a very significant interaction effect between the two tap water attributes. Since the attributes describe days of tap water problems, the negatively signed interaction effect can be interpreted as suggesting that the attributes have a substitute relationship; respondents attach less value to improvements in one tap water attribute the more days of the other improvement they are offered. Where the interaction of COST with a water quality attribute is significant the parameter is positively signed indicating that respondents become increasingly price sensitive at high package costs.[10]

[10] One explanation of that observation is that an increasingly large proportion of the sample encounters budget constraints as the package costs rise. Consider a respondent who regards an option offering moderate water improvements at a moderate cost as being preferable to the SQ. Imagine another option that scales-up the levels of water improvements and cost of the original option by a fixed proportion. If preferences can be perfectly described by the main effects of the attribute levels then the respondent would necessarily also choose the scaled-up option over the SQ. Of course, for some respondents the increased cost of the scaled-up option may be unaffordable given the constraint imposed by the budget from which this cost must be

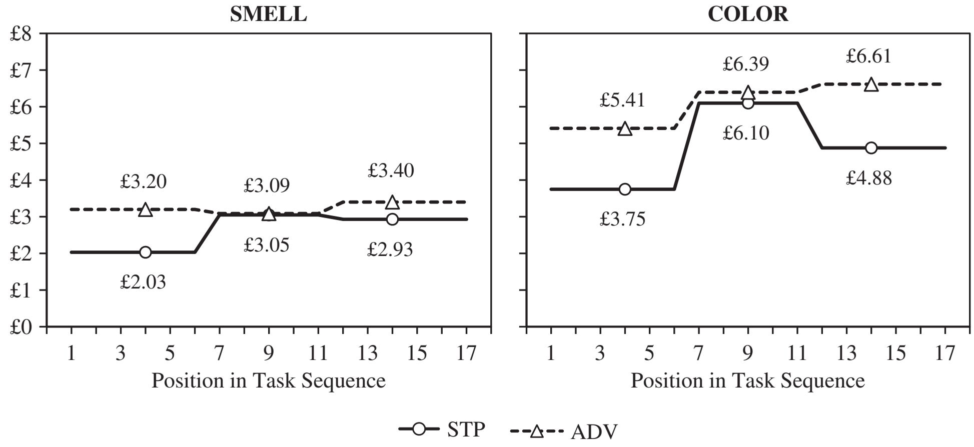

Fig. 6. Position dependence in marginal WTP for tap water quality improvements.

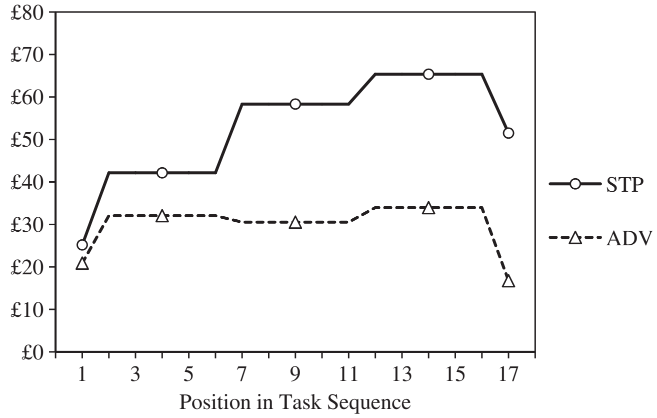

Fig. 7. Position dependence in the location parameter of the status quo effect (in the WTP metric).

6.3. Status quo effect

In both the STP and ADV models the SQ effect coefficient, α, is negative and statistically highly significant. It appears that the mean behavior in both samples can be characterized as a tendency to reject the alternative option in favor of the SQ, independent of the quality of the deal offered by that alternative option. Once again, however, our models reveal a significant difference between the two treatment groups regarding how the SQ effect develops over the sequence of tasks. To facilitate comparison, the development of the SQ effect for each group is plotted out in the WTP metric in Fig. 7. These plots can be thought of as detailing each sample's aversion to choosing the costly option offered as an alternative to the SQ; the greater the value of the SQ effect the larger that aversion.

In the first task the SQ effects of the STP group and ADV group are relatively similar. Indeed we cannot reject the hypothesis that they take the same value (p-value: 0.2514). For both groups, the SQ effect increases significantly moving into the next five tasks. That increased aversion to the alternative option is most marked in the STP group such that over the second to sixth tasks the SQ effect for the STP group exceeds that of the ADV group by an amount that is statistically significant at the 90% level of confidence (p-value: 0.0994). For the ADV group the SQ effect remains statistically constant across the middle 15 tasks of the design, a pattern in stark contrast to that exhibited by the STP group. Over those same tasks, we observe a progressive increase in the STP group's aversion to the alternative option, such that over tasks 7–16 the SQ effect of the STP group is larger than that of the ADV group by an amount that is statistically significant at greater than the 95% level of confidence (tasks 7–11, p-value: 0.0089; tasks 12–16, p-value: 0.0105). In both groups, the SQ effect falls in the final task. For the ADV group, that final downturn results in a SQ effect in responses to the final task that is statistically

(*footnote continued*)
met. Accordingly, binding budget constraints will lead to increased levels of rejection of options at higher costs even if those options are associated with very large tap water quality improvements.

indistinguishable from the SQ effect in responses to the identical opening task (p-value: 0.7377). In the STP group, on the other hand, aversion to the alternative option is significantly higher in the final task than it is in the first task (p-value: 0.0009).

Fig. 7 suggests that respondents treat the first and last task somewhat differently to the intervening tasks. We must be cautious, however, in attributing that finding entirely to an SQ effect. Since our design precludes the estimation of attribute taste parameters specific to the first and last tasks, the observed pattern could also be attributable to differences in preferences for the COST, COLOR or SMELL attributes in those tasks. Either way, it is apparent that in both groups there are important behavioral differences in responses to the repeated task.

6.4. Precedent dependent effects

To examine the possibility that choices are framed by comparison to options in preceding tasks, our model includes the values of the deals offered in four different preceding tasks; the deals in the first and last tasks, plus the worst and best deals seen so far. For both treatment groups, the value of the deal in the first task plays an important role in shaping preferences, as does the value of the best deal seen so far. In contrast the deal observed in the preceding task has no significant impact on the preferences of either the STP or ADV group while the worst deal seen so far is important in shaping the preferences of the STP group but not the ADV group. In all cases, parameters of significant precedent-dependent effects are positive, indicating that the deal in the current task will be considered more favorably if it exceeds the precedent set by that previously observed deal than if it falls short of that mark.

Our modeling specification allows us to conceptualize the combined influence of preceding options in the form of a *reference deal* formed through the linear combination of particular previous deals. Using (7), and assuming insignificant parameters to take a value of zero, our model suggests that the reference deal of the STP group is the weighted average of the start deal, worst deal and best deal, applying weights of 35%, 52% and 13% respectively. In contrast, the ADV group pays no attention to the worst deal. For this group, the reference deal is a linear combination of the start deal and best deal, using weights of 55% and 45%, respectively.

Our specification also allows us to examine the strength of the referencing effect (given by the sum of precedent-dependent parameters; $g=\gamma_1+\gamma_2+\gamma_3+\gamma_4$); that is to say, to measure the extent to which comparison to the reference deal impacts on the evaluation of the deal in the current task. We find that g takes a value of 0.72 for the ADV group and 0.99 for the STP group, though the difference between the two is not significant (p-value: 0.2226).

Accordingly, while both groups exhibit marked referencing behavior, it appears that the method of task revelation is influential in determining how respondents form their reference deal. While the value of the start deal is important to both groups, the STP group is strongly influenced by the worst deal whereas the ADV group focuses on the best deal.

6.5. Error scale

The final set of parameters in Table 2 report the scale of the error distribution and its development over the sequence of tasks. Recall that to allow for individual heterogeneity the error scale is treated as a random parameter with a log-normal distribution, a specification supported by the fact that the estimated scale of that assumed distribution, s_σ, is highly significant in both models. Note that on their own, the position dependence parameters relating to the error scale are rather difficult to interpret because they describe shifts in the location parameter of a log-normal distribution. To facilitate interpretation, therefore, Fig. 8 plots out the mean of the error scale distribution across the sequence of tasks.

Across the middle 15 tasks the error scale for both groups remains relatively stable; there are no statistically significant changes for the STP group and, for the ADV group, the reduction in the error scale over tasks eleven to sixteen is only

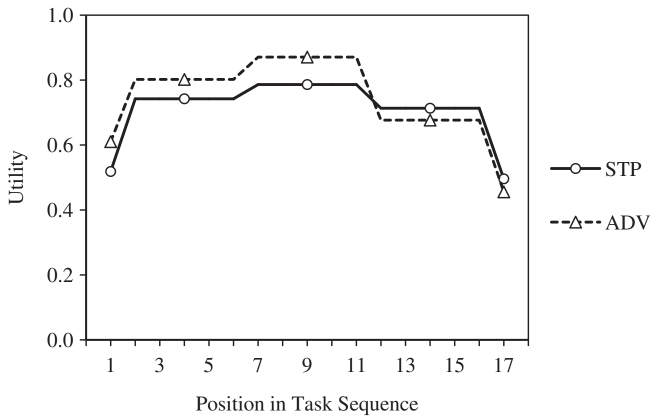

Fig. 8. Position-dependence in the mean of the error scale distribution (in the utility metric).

significant at the 90% confidence level (*p*-value: 0.0675). Both models also suggest a relatively smaller error scale in responses to the first and last task. As discussed in Section 3.2, however, that apparent difference in scale may simply reflect differences in the experimental design in those task positions. As such, we refrain from affording that observation behavioral significance. Finally, we find that in all positions in the sequence of tasks there is no significant difference in the error scale across the two treatments. Accordingly, our models provide little evidence to suggest the error scale changes systematically along the sequence of tasks or differs between STP and ADV treatments.

7. Discussion

Both our non-parametric and parametric analyses support the same conclusion; choices made in this DCE exhibit a variety of both position- and precedent-dependent order effects. Let us now turn to consider how those observed patterns of order effect compare to the predictions made by the various competing theories reviewed in Section 2.

One striking finding of our research is that we observe significant position dependence in the preferences of the STP group but not the ADV group. The exact pattern of that position dependence provides insights as to why that may be so. First, as they progress through the tasks the STP group's valuation of attributes converges on the stable values expressed by the ADV group. That pattern is commensurate with the theory of *institutional learning*, but not of *preference learning*. In particular, preference learning is reputedly achieved through the experience of trading-off attributes when making decisions in choice tasks. Since that decision-making experience is identical for STP and ADV groups, preference learning predicts that preferences will change in an identical manner for both groups. In contrast, working through the tasks, the STP group's understanding of the nature of the DCE exercise grows to match that endowed to the ADV group at the outset. Institutional learning, therefore, provides a passable explanation of the pattern observed in our data.

The second pattern of position dependence in the STP group's choices concerns their predilection for choosing the SQ option. While initially similar to that of the ADV group, progressing through the exercise respondents in the STP group show an increasing tendency to select the SQ option over alternative options. That pattern is anticipated by the *failing credibility* hypothesis but, as discussed shortly, not by other theories. In particular, since the ADV group is aware of the full sequence of tasks from the outset, there is no reason to suspect that their attitudes towards the alternative options will change over the course of the exercise. This is what we observe in our data. In contrast, for the STP group each successive task is unannounced and possibly unanticipated. The failing credibility hypothesis conjectures that the unannounced presentation of differently configured alternative options progressively undermines respondents' willingness to accept those options as outcomes of genuine programs of work increasing their propensity to reject them in favor of the SQ. Again this is what we observe in our data.

While the failing credibility hypothesis predicts the rising SQ effect over the first 16 tasks, it does not accommodate the apparently anomalous fall in the SQ effect in the final task. One possible explanation follows from the fact that the final task is identical to the first. If respondents are motivated by a desire to answer in an internally consistent manner, a behavior well documented in experimental economics [2,3], then they may choose to answer the final task in the same way as they did the first.

An alternative explanation for an increasing SQ effect is that respondents become progressively *fatigued* by the cognitive burden of assessing the deals offered by the alternative options and, as they work through the tasks, increasingly opt for the "safe choice" offered by the SQ. Of course, we have no reason to suspect that fatigue would more greatly afflict respondents in the STP group than the ADV group, such that the patterns of behavior observed in our data do not sit easily with a fatigue explanation. Moreover, recall that we find no evidence of increasing variability in respondents' answers as they progress through the tasks, another pattern that has been mooted as indicative of fatigue. All in all, our data provide little evidence to suggest that fatigue effects are driving position-dependent order effects.

The income uncertainty hypothesis may also explain position-dependence in choices. That theory predicts that WTP may fall along the sequence of tasks as the presentation of differently costed options progressively alerts respondents in the STP group to potential variance in their future income. In accordance with that prediction, our model suggests a position-dependent fall in WTP for the STP group, a fall that is driven by an increasing proclivity to choose the SQ option (and not through increasing cost sensitivity or a falling assessment of the value of water quality attributes). In contrast, our model does not identify position dependence as driving a fall in WTP for the ADV group such that the patterns observed in our data are also commensurate with the income uncertainty hypothesis.[11]

As well as position dependence in preferences, our data provides robust evidence of precedence dependence. Interestingly, this appears to be true of both treatment groups; precedent-dependent order effects, unlike position-dependent order effects, are not mitigated by advance disclosure of tasks. Both groups, for example, exhibit a strong starting point effect. Starting point phenomenon are well documented in SP studies (e.g. [10,27,36]) where the most widely posited explanation is that uncertainty and unfamiliarity lead respondents to *anchor* their preferences to features of the first preference elicitation task. That explanation does not sit entirely comfortably with our findings; one might

[11] The data are not wholly supportive of the income uncertainty model. Unless income uncertainty is substantially larger in the STP group, one might expect the ADV group (to whom the uncertainty in income becomes evident at the start of the process) to exhibit lower WTP in the initial task than the STP group (to whom that uncertainty is only subsequently revealed). Our nonparametric and parametric analyses find no evidence of such a difference.

justifiably expect that advanced revelation of the features of future tasks would reduce respondents' propensity to anchor on the particular features of the first task, a pattern that we do not observe. The strong starting point effect in the ADV group suggests that this may be a recalcitrant choice heuristic in SP exercises.[12]

Our analysis also reveals that respondents' choices are influenced by the best and worst deals that they have been presented with previously in the exercise. Accordingly, it appears that precedence dependence does not necessarily work by simply fixing individuals preferences to the features of, say, the first task. Rather, individuals' preferences may develop in a dynamic way being shaped both by the first task and the changing information they receive on the range of possible deals. That interpretation closely resembles the *reference effect hypothesis* which posits that individuals judge options as representing good or bad deals relative to a reference deal developed from past experience of possible deals.

Our modeling specification is amenable to interpretation through the lens of referencing behavior. In particular, we are able to estimate the implied reference deal as a linear combination of particular previous deals. From this perspective, the reference deals of both treatment groups are seen to be strongly influenced by the deal offered in the arbitrarily attributed starting task but not by the deal offered in the immediately preceding task. The two treatment groups show very different behavior, however, with respect to how the reference deal is shaped by the best and worst deals seen so far. Respondents in the STP group form a reference deal focused towards the bottom of the range between those two deals, while respondents in the ADV group form a reference deal focused on the very top of that range.

The standard conceptualization of the reference deal is that it is formed as an expectation based on the quality of previously offered deals [40]. Accordingly, the marked asymmetry between STP and ADV groups could be interpreted as indicating different risk attitudes in the formation of that expectation. Respondents in the STP group, unaware of the number and nature of future choice tasks, adopt a cautious perspective in assessing what constitutes a good deal. Respondents in the ADV group, aware of the full range of tasks, are more cavalier. An alternative explanation might be found in incentives for *strategic misrepresentation*. Advanced revelation of tasks alerts respondents to the existence of multiple-possible deals. Respondents might interpret that format as indicating that the exercise is for the purposes of deciding which particular deal should be offered in reality. In those circumstances, one obviously beneficial strategy would be to favor options offering deals better than the best deal seen so far whilst underplaying the strength of one's preferences for options that fall short of that mark.

Finally, recall that our non-parametric analysis reveals that both treatment groups exhibit a significant shift in preferences between the first and last task. Our parametric analysis suggests that the reasons for that shift may differ across treatments. In particular, the stepwise revelation of tasks appears to invoke position dependence in preferences. Indeed, the observed change in preferences between first and last task for that group appears to result primarily from a progressive shift towards favoring the SQ option. In contrast, the ADV disclosure of tasks forewarns individuals of the repeated nature of the exercise and that information appears to greatly mitigate position-dependent order effects. While both groups exhibit clear evidence of precedent-dependent preferences, the ADV group, in contrast to the STP group, form a reference deal focused on the best deal seen previously. Of course, as the ADV group progress through the tasks, the best deal they have been offered can only increase in value. It follows that other deals, when compared to that increasing reference, are cast in a progressively less positive light. Indeed, it is this effect that our models suggest primarily explains the shift in the ADV group's preferences between the first and last task.

8. Concluding remarks

The central contribution of this research has been to outline a design and modeling framework that allows for the simultaneous investigation of a wide variety of order effects in a DCE. Central to that methodology is a comparison of advanced and stepwise disclosure formats. In our empirical application that comparison proves a powerful source of identification for distinguishing between proposed theories of order effect phenomena.

The first key finding of our empirical application is that position-dependent order effects are related primarily to the stepwise revelation of tasks, perhaps precipitated by institutional learning and failing credibility or maybe by increasing income uncertainty. By familiarizing individuals with the exercise format, the advanced disclosure of tasks appears to largely mitigate those effects. That is not to say that advanced disclosure provides a cure-all for order effects in repeat-response SP exercises, far from it. Our second key finding is that precedent-dependent order effects pervade the choices of respondents in both revelation formats. Once again, however, differences between the STP and ADV treatment groups suggest that the expression of those effects may be shaped by the nature of task revelation, possibly through the encouragement of strategic misrepresentation of preferences in the latter group's choices.

Since our analyses provide robust evidence of order effects, a central message of this research is that our data do not support a naïve interpretation of choices in repeat-response SP exercises as revealing information on some notion of fixed and stable 'true' preferences. Importantly, that finding should not necessarily be taken as evidence of some inherent problem with the elicitation of preferences using repeat-response SP formats. For example, precedent-dependent order effects similar to those observed in our data have also been observed in real purchasing decisions (e.g. [32,45,40]) and in

[12] Ariely et al. [3] report on an experiment in which participants exhibit significant starting point effects despite being made explicitly aware of the existence of multiple possible starting points.

incentivized experimental markets (e.g. [32]). Such patterns of response may, therefore, be symptomatic of some more general process of preference formation and expression; a proposition requiring further research.

Nevertheless, one might conclude that the presence of precedent-dependent order effects makes it difficult to draw reliable inference from responses to any but the initial task. By virtue of its position in the sequence, of course, that task cannot be subject to such effects. With regards to the properties of responses in that initial task, our experiment is also informative. Our non-parametric analysis shows this task has excellent properties with respect to the standard criteria of price and scope sensitivity. Moreover, we find that the SQ effect in the initial task is similar under both the ADV and STP treatments and significantly lower than in later tasks. That finding is also encouraging insomuch as it suggests that in the initial task respondents are more greatly inclined to judge options solely on the merits of their attributes.[13] At the same time, our analyses indicate that in the initial task the STP group express relatively lower values for attributes than the ADV group. Since the values of the STP group subsequently converge on those of the ADV group, their initially lower values may relate to a lack of understanding of, or familiarity with the institution of preference expression in a DCE exercise. If that interpretation of our data proves robust, then it would suggest that task training and information provision may be required to mitigate order effects in the initial tasks of a DCE.

Of course, an alternative perspective would be that the initial differences between the ADV and STP treatments are actually precipitated by the additional information provided to the ADV group. In particular, the ADV group's opening decisions may be influenced by their knowledge of options in subsequent tasks. To extend our terminology, ADV revelation admits the possibility of *subsequent-dependent order effects*. Our experiment does not allow us to say a great deal about that possibility. Specifically, the best and worst deals in the subsequent tasks are initially identical for all ADV respondents such that we cannot identify whether the nature of those deals has any bearing on their initial choices. Likewise, in our design the last task is identical to the first task, so a subsequent-dependent *last-task effect* cannot be independently identified from the precedent-dependent *first-task effect*.

Our experiment, therefore, suggests that a potentially fruitful area of further research will be to examine how the preferences expressed in DCE exercises respond to different types of information provision (spanning the range between the minimal information of our STP treatment to the complete information of our ADV treatment) and different forms of training in decision-making in the choice task format. The goal of that research would be to identify ways to structure DCEs that reduce the position-dependent order effects (as observed in our STP group) without inducing undesired subsequent-dependent referencing behavior. Such a research program could also shed further light on the ways in which information provision alters precedent-dependent referencing behavior (as observed in the differences in that behavior between our STP and ADV groups).

While we make no claims regarding the generality of our findings (our empirical results are particular to the data set we have studied) the research reported in this paper provides some intriguing insights into the nature and possible causes of order effects in DCEs. We hope, in addition, that the experimental design and modeling framework described here provide useful contributions to the on-going study of those phenomena.

Acknowledgments

We are very grateful to two anonymous referees and the Editor for helpful comments on initial drafts of this paper. This research was supported by the ESRC through the Social and Environmental Economic Research (SEER) project (Ref RES-060-25-0063). Support from the Economics for the Environment Consultancy (eftec) is also gratefully acknowledged.

References

[1] W.L. Adamowicz, What's it worth? An examination of historical trends and future directions in environmental valuation, Australian Journal of Agricultural and Resource Economics 48 (2004) 419–443.
[2] D. Ariely, G. Loewenstein, D. Prelec, Coherent arbitrariness: stable demand curves without stable preferences, Quarterly Journal of Economics 118 (2003) 73–105.
[3] D. Ariely, G. Loewenstein, D. Prelec, Tom Sawyer and the construction of value, Journal of Economic Behavior & Organization 60 (2006) 1–10.
[4] I.J. Bateman, D. Burgess, W.G. Hutchinson, D.I. Matthews, Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness, Journal of Environmental Economics and Management 55 (2008) 127–141.
[5] I.J. Bateman, M. Cole, P. Cooper, S. Georgiou, D. Hadley, G.L. Poe, On visible choice sets and scope sensitivity, Journal of Environmental Economics and Management 47 (2004) 71–93.
[6] I.J. Bateman, B.H. Day, D.P. Dupont, S. Georgiou, Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method, Review of Economics and Statistics 91 (2009) 806–820.
[7] I.J. Bateman, B.H. Day, A.P. Jones, S. Jude, Reducing gain-loss asymmetry: a virtual reality choice experiment valuing land use change, Journal of Environmental Economics and Management 58 (2009) 106–118.
[8] I.J. Bateman, I.H. Langford, A.P. Jones, G.N. Kerr, Bound and path effects in double and triple bounded dichotomous choice contingent valuation, Resource and Energy Economics 23 (2001) 191–213.
[9] I.J. Bateman, A. Munro, G.L. Poe, Decoy effects in choice experiments and contingent valuation: asymmetric dominance, Land Economics 84 (2008) 115–127.

[13] One way to correct for order effects in which respondents increasingly favor the SQ option would be to normalize the SQ effect in the estimated preference function to this initial level.

[10] K.J. Boyle, R.C. Bishop, M.P. Welsh, Starting point bias in contingent valuation bidding games, Land Economics 61 (1985) 188–194.
[11] M. Bradley, A. Daly, Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data, Transportation 21 (1994) 167–184.
[12] J. Braga, C. Starmer, Preference anomalies, preference elicitation and the discovered preference hypothesis, Environmental & Resource Economics 32 (2005) 55–89.
[13] J.D. Brazell, J. Gray-Lee, J.J. Louviere, Revisting order effects: are they just differences in random variation?, in: Proceedings of the Marketing Science Conference, Fontainbleu, France, 1998.
[14] J.D. Brazell, J. Gray-Lee, J.J. Louviere, B. Dallaert, P. Pullman, Nobody will do that! An empirical investigation of survey length effects on response rate and reliability in choice model experiments, in: Proceedings of the INFORMS Marketing Science Conference, Sydney Australia, 1995.
[15] J.S. Butler, R. Moffitt, A computationally efficient quadrature procedure for the one-factor multinomial probit model, Econometrica 50 (1982) 761–764.
[16] T.A. Cameron, J. Quiggin, Estimation using contingent valuation data from a "dichotomous choice with follow-up" questionnaire, Journal of Environmental Economics and Management 27 (1994) 218–234.
[17] R.T. Carson, T. Groves, Incentive and informational properties of preference questions, Environmental and Resource Economics 37 (2007) 181–210.
[18] J. Clark, L. Friesen, The causes of order effects in contingent valuation surveys: an experimental investigation, Journal of Environmental Economics and Management 56 (2008) 195–206.
[19] B. Day, J.L.P. Prades, Ordering anomalies in choice experiments, Journal of Environmental Economics and Management 59 (2010) 271–285.
[20] W.S. DeSarbo, D.R. Lehmann, F.G. Hollman, Modeling dynamic effects in repeated-measures experiments involving preference/choice: an illustration involving stated preference analysis, Applied Psychological Measurement 28 (2004) 186–209.
[21] J.R. DeShazo, Designing transactions without framing effects in iterative question formats, Journal of Environmental Economics and Management 43 (2002) 360–385.
[22] J.R. DeShazo, G. Fermo, Designing choice sets for stated preference methods: the effects of complexity on choice consistency, Journal of Environmental Economics and Management 44 (2002) 123–143.
[23] V. Foster, S. Mourato, Testing for consistency in contingent ranking experiments, Journal of Environmental Economics and Management 44 (2002) 309–328.
[24] M. Hanemann, J. Loomis, B. Kanninen, Statistical efficiency of double-bounded dichotomous choice contingent valuation, American Journal of Agricultural Economics 73 (1991) 1255–1263.
[25] J.J. Heckman, Statistical models for discrete panel data, in: C.F. Manski, D. McFadden (Eds.), Structural Analysis of Discrete Data and Econometric Applications, MIT Press, Cambridge, 1981, pp. 114–178.
[26] D. Hensher, W. Greene, Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification, Empirical Economics 39 (2010) 413–426.
[27] J.A. Herriges, J.F. Shogren, Starting point bias in dichotomous choice valuation with follow-up questioning, Journal of Environmental Economics and Management 30 (1996) 112–131.
[28] R.M. Hogarth, H.J. Einhorn, Order effects in belief updating: the belief-adjustment model, Cognitive Psychology 24 (1992) 1–55.
[29] T.P. Holmes, K.J. Boyle, Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions, Land Economics 81 (2005) 114–126.
[30] D. Hoyos, The state of the art of environmental valuation with discrete choice experiments, Ecological Economics 69 (2010) 1595–1603.
[31] E.R. Igou, H. Bless, Inferring the importance of arguments: order effects and conversational rules, Journal of Experimental Social Psychology 39 (2003) 91–99.
[32] A. Isoni, The willingness-to-accept/willingness-to-pay disparity in repeated markets: loss aversion or 'bad-deal' aversion? Theory and Decision 71 (2011) 409–430.
[33] F.R. Johnson, W.H. Desvousges, Estimating stated preferences with rated-pair data: environmental, health, and employment effects of energy programs, Journal of Environmental Economics and Management 34 (1997) 79–99.
[34] B.J. Kanninen, Optimal design for multinomial choice experiments, Journal of Marketing Research (JMR) 39 (2002) 214–227.
[35] W.F. Kuhfeld, R.D. Tobias, M. Garratt, Efficient experimental design with marketing research applications, Journal of Marketing Research (JMR) 31 (1994) 545–557.
[36] J. Ladenburg, S.B. Olsen, Gender-specific starting point bias in choice experiments: evidence from an empirical study, Journal of Environmental Economics and Management 56 (2008) 275–285.
[37] K.J. Lancaster, New approach to consumer theory, Journal of Political Economy 74 (1966) 132–157.
[38] J.A. List, Does market experience eliminate market anomalies? Quarterly Journal of Economics 118 (2003) 41–71.
[39] J.J. Louviere, D.A. Hensher, J.D. Swait, Stated Choice Methods : Analysis and Application, Cambridge University Press, Cambridge, 2000.
[40] T. Mazumdar, S.P. Raj, I. Sinha, Reference price research: review and propositions, Journal of Marketing 69 (2005) 84–102.
[41] D. McFadden, Conditional logit analysis of qualitative choice behavior, in: P. Zarembka (Ed.), Frontiers in Econometrics, Academic Press, New York, 1973.
[42] MORI, The 2004 periodic review: research into customers' views, London, 2002.
[43] T. Ohler, A. Le, J. Louviere, J. Swait, Attribute range effects in binary response tasks, Marketing Letters 11 (2000) 249–260.
[44] C.R. Plott, Rational individual behavior in markets and social choice processes: the discovered preference hypothesis, in: K. Arrow, M. Colombatto, M. Perleman, C. Schmidt (Eds.), Rational Foundations of Economic Behavior, MacMillan, London, 1996, pp. 225–250.
[45] C.R. Plott, K. Zeiler, The willingness to pay-willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations, American Economic Review 95 (2005) 530–545.
[46] G.L. Poe, E.K. Severance-Lossin, M.P. Welsh, Measuring the difference ($X-Y$) of simulated distributions: a convolutions approach, American Journal of Agricultural Economics 76 (1994) 904–915.
[47] R.C. Ready, J.C. Whitehead, G.C. Blomquist, Contingent valuation when respondents are ambivalent, Journal of Environmental Economics and Management 29 (1995) 181–196.
[48] S.J. Savage, D.M. Waldman, Learning and fatigue during choice experiments: a comparison of online and mail survey modes, Journal of Applied Econometrics 23 (2008) 351–371.
[49] R. Scarpa, S. Notaro, J. Louviere, R. Raffaelli, Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons, American Journal of Agricultural Economics 93 (2011) 809–824.
[50] A.P. Street, D.J. Street, Combinatorics of Experimental Design, Oxford University Press, Oxford, 1986.
[51] J. Swait, W. Adamowicz, The influence of task complexity on consumer choice: a latent class model of decision strategy switching, Journal of Consumer Research 28 (2001) 135–148.
[52] A. Tversky, Elimination by aspects: a theory of choice, Psychological Review 79 (1972) 281–299.
[53] G.C. van Kooten, E. Krcmar, E.H. Bulte, Preference uncertainty in non-market valuation: a fuzzy approach, American Journal of Agricultural Economics 83 (2001) 487–500.
[54] P.H. Wicksteed, The Common Sense of Political Economy, Routledge, London, 1910.