

Day, Brett et al.

Working Paper

Task independence in stated preference studies: A test of order effect explanations

CSERGE Working Paper EDM, No. 09-14

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Day, Brett et al. (2009) : Task independence in stated preference studies: A test of order effect explanations, CSERGE Working Paper EDM, No. 09-14, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/48813>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Task Independence in Stated Preference Studies:
A Test of Order Effect Explanations**

by

Brett Day, Ian Bateman, Richard Carson, Diane
Dupont, Jordan Louviere, Sanae Morimoto, Sanae
Morimoto, Riccardo Scarpa and Paul Wang

CSERGE Working Paper EDM 09-14

**Task Independence in Stated Preference Studies:
A Test of Order Effect Explanations**

30th November 2009

Brett Day
University of East Anglia, UK

Ian J. Bateman
University of East Anglia, UK

Richard T. Carson
University of California, San Diego, USA

Diane Dupont
Brock University, Canada

Jordan J. Louviere
University of Technology, Sydney, Australia

Sanae Morimoto
Kobe University, Japan

Riccardo Scarpa
University of Waikato, New Zealand

Paul Wang
University of Technology, Sydney, Australia

Address for manuscript correspondence:

Professor Ian J. Bateman
Centre for Social and Economic Research on the Global Environment
School of Environmental Sciences
University of East Anglia
Norwich
NR4 7TJ
England
United Kingdom

Tel: 00 44 (0) 1603 593125
Fax: 00 44 (0) 1603 593739
email: i.bateman@uea.ac.uk

Task Independence in Stated Preference Studies: A Test of Order Effect Explanations¹

Abstract

We present an experiment designed to investigate the presence and nature of ordering effects within repeated response stated preference studies. We formulate a general structural model of such effects and use this to isolate signature patterns for position-dependent effects (learning about preferences or institutions, and the impact of fatigue) and precedent-dependent effects (starting point effects, reference pricing and various forms of strategic behavior). This is tested within a large sample, full factorial study designed to mitigate against misspecification bias and design-induced error variance problems. Non-parametric and parametric analyses are applied, the latter adopting a novel data-driven approach to the detection of ordering patterns. While we find little evidence of position dependent effects, we do find evidence of a starting point effect and various types of strategic behavior including a reference price effect where respondents tend to reject alternatives that are priced higher than recently seen alternatives.

Key words

Stated preference, ordering effect, strategic behavior, reference price, drinking water

¹ Acknowledgements: We are very grateful to two anonymous referees and the Editor for helpful comments on an initial draft 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.

1. Introduction

As measured by either refereed journal papers, textbooks or official research reports, stated preference (SP) methods now play a major role in empirical efforts to place monetary values upon preferences for non-market goods such as those delivered by the natural environment (Carson, forthcoming; Champ et al., 2003; Defra, 2007). While the Arrow, et al. (1993) report on contingent valuation help enshrine the single binary discrete choice (SBC) variant of contingent valuation (CV) as the gold standard, the statistical inefficiency of this approach led to the development of a range of more efficient, repeated response variants (e.g. Hanemann et al., 1991). This trend is extended through the more recent rise in the popularity of more general discrete choice experiment (DCE) approaches for which repeated choices are a fundamental element of study design (Adamowicz, et al., 1994; Louviere, et al., 2000; Adamowicz, 2004).

In the move from single to repeated response formats the assumption is typically made that respondents “regard each valuation task as independent and exclusive of all other tasks” (Day and Pinto, 2009). However, the literature has reported violations of the independence assumption (Cameron and Quiggin, 1994; Herriges and Shogren, 1996; Alberini, Kanninen and Carson, 1997; Johnson and Desvousges, 1997; Bateman, et al., 2001; Swait and Adamowicz, 2001; DeShazo, 2002; Holmes and Boyle, 2005). Although these studies propose a variety of differing drivers underpinning these violations of the independence assumption, the common effect uniting these findings is the observation of an impact of question ordering upon responses.

The present paper seeks to contribute to this literature by investigating the nature and motivations underpinning such ordering effects. Only by understanding such motivations can we assess the validity of responses exhibiting such effects and determine the usefulness of resultant valuations for decision making purposes. Our study opens in Section 2 with a consideration of the standard structural model of choice underpinning most SP studies. We employ and extend this model to formally identify signature response patterns for discriminating between a variety of drivers of ordering effects which we find in our literature reported in Section 3. It is these signature patterns which we test for in our subsequent empirical analysis of ordering effects.

Section 4 describes the design of a study to analyze these signature response patterns and so discriminate between competing explanations of the nature and interpretation of ordering effects. This empirical investigation utilizes a repeated choice SP study executed using a large sample survey, implemented using high quality, face-to-face interviewing techniques and concerning domestic tap water supplies described via two quality attributes and a cost difference. The study adopts a Latin square approach to the allocation of choices into a question

sequence so as to mitigate against design induced ordering effects while a full factorial of options is applied to avoid misspecification bias through the incorporation of interaction terms. Further aspects of the design permit investigation of competing explanations of ordering effects using both robust non-parametric tests and parametric analyses, the latter adopting a novel data-driven approach to the detection of ordering patterns. Section 5 sets out our approach to modeling while results are reported in Section 6. These confirm the presence of more than one form of ordering effect within our data. Section 7 provides concluding observations on what we believe has been learned from this exercise and its potential usefulness in future applications.

2. Structural model of responses to a choice based SP exercise

The fundamental unit of choice based SP exercise is an alternative or option. An option presents the respondent with a possible level of provision of the non-market good, usually defined by the levels of certain attributes of that good, along with an associated cost. The most basic format is an SBC between two alternatives. This format can be extended by moving from a binary to a multinomial choice question (the response to which can always be decomposed into a set of implied SBC comparisons). The format can also be extended by asking a sequence of either binary choice questions, which is often called a sequence of paired comparisons or a sequence of multinomial choice questions. The standard structural model of choice for such tasks is provided by McFadden's (1974) familiar random utility framework. Here, respondents are assumed to accord a level of utility to an option based, in part, on the levels of attributes of the good on offer and its associated cost. In addition, the structural model includes a random element to utility that is usually motivated as the inability of the analyst to observe all the relevant attributes of the choice task. More formally, imagine a set of options, denoted \mathbb{A} , constructed for the purpose of the DCE. Also let ω and ϖ denote particular options in that set. Options are selected (often with replacement) from \mathbb{A} for inclusion in one of a series of choice tasks indexed $j = 1, 2, \dots, J$. We denote by $\mathbb{A}_j \subset \mathbb{A}$ the set of two or more options in choice task j .

Now assume that the DCE is administered to a sample of respondents indexed $i = 1, 2, \dots, N$. Then, the usual point of departure, is to assume that the utility derived by respondent i from a particular option, ω , when presented in choice task j can be approximated by a linear utility function of the form;

$$U_{\omega,ij} = \alpha_{\omega,i} + \mathbf{X}'_{\omega} \boldsymbol{\beta}_i + e_{\omega,ij} \quad (1)$$

where;

\mathbf{X}_{ω} is the M -dimensional vector of cost and attribute levels for option ω .

$\alpha_{\mathbb{O},i}$ is the parameter capturing any utility respondent i derives from choosing option \mathbb{O} independent of the attributes of that option.

$\boldsymbol{\beta}_i$ is the M -dimensional vector of parameters capturing respondent i 's marginal (dis)utilities of attributes and costs.

$e_{\mathbb{O},ij}$ is the random element usually assumed to be i.i.d. with variance σ^2 .

According to this standard presentation, respondent i is assumed to prefer option \mathbb{O} when presented in choice task j if;

$$U_{\mathbb{O},ij} > U_{\mathbb{Q},ij} \quad (\text{for all } \mathbb{Q} \in \mathbb{A}_j; \mathbb{Q} \neq \mathbb{O}) \quad (2)$$

Observe that the specification in (1) suggests that the utility of a particular option is independent of;

- where in the sequence of choice tasks, the option is presented.
- the nature of options presented in previous choice tasks;
- the nature of the other options presented in that choice task;

This paper is motivated by a growing body of evidence (reviewed in Section 3) suggesting that these three independence assumptions may not always hold. The particular focus of this study is on violations of the first two assumptions in the list as, taken together, such violations manifest themselves as the most frequently observed anomaly in repeated response SP studies to date; *order effects* (to avoid the possibility of other options in a choice task affecting responses to an option we use a simple design with a constant ‘status quo’ option being compared with a changing alternative option). Such effects manifest themselves as the observation that choice behavior changes systematically over the course of a series of SP tasks. Drawing on the nomenclature of Day and Pinto (2009), violations of independence from position in the sequence of tasks result in what can be termed *position-dependent* order effects, while violations of independence from the nature of options in previous choice tasks result in what can be termed *precedent-dependent* order effects.

In terms of the model in (1), position-dependent order effects could be captured through a revision of the utility specification of the form;

$$U_{\mathbb{O},ij} = \alpha(j)_{\mathbb{O},i} + \mathbf{X}'_{\mathbb{O}} \boldsymbol{\beta}(j)_i + e(j)_{\mathbb{O},ij} \quad (3)$$

where $\beta(j)_i$, $\alpha(j)_{\circ,i}$ and $e(j)_{\circ,ij}$ imply that the preference parameters and error term depend in some way on position in the sequence of tasks.

Likewise, a possible way to introduce precedent-dependent order effects into the specification is through the inclusion of one or more *precedent variables*;

$$U_{\circ,ij} = \alpha(j)_{\circ,i} + \mathbf{X}'_{\circ} \beta(j)_i + \mathbf{P}(\mathbf{X}_{\circ}; \mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1})' \boldsymbol{\theta}_i + e(j)_{\circ,ij} \quad (4)$$

The vector of precedent variables, $\mathbf{P}(\mathbf{X}_{\circ}; \mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1})$, captures the nature of the comparisons made by a respondent between the attributes of the current option, \circ , when presented in choice task j and the attributes of other options observed. The impact of the various precedent variables on the perceived utility of option \circ in choice task j is captured through the vector of parameters $\boldsymbol{\theta}_i$.

Notice that the order effects that form the focus of our study involve precedent variables of the form ; that is to say, where that compare the attributes of the current option to those of options in preceding tasks. Our notation, however, is general enough to allow for precedent variables of the form $\mathbf{P}(\mathbf{X}_{\circ}; \mathbf{X}_{A_j})$; that is to say, where the utility of option \circ is somehow affected by comparison with only the options in that same choice task. Precedent variables of that description would capture violations of the final independence assumption listed above. As we discuss subsequently, so as to focus on the issue of order effects we design an experiment that eliminates the possibility of violations of that sort.

3. Literature review and identification of signature patterns for ordering effects

Recent years have seen an increase in the number of studies giving consideration to the causes of ordering effects in choices made in sequential SP tasks. This is supplemented by a longer established literature in industrial organization and marketing considering the same subject in the context of purchasing behavior. While these literatures may have differing objectives (the assessment of welfare values as opposed to the extraction of producer surplus) both offer insights into the nature of ordering effects. A review of both literatures identified six alternative accounts of ordering effects (none of which are necessarily mutually exclusive), three of which can be categorized as position-dependent (preference learning, institutional learning and fatigue) while the remainder are precedent dependent (starting point effects, and two variants of strategic behavior that differ by the amount of price recall, one of which we will refer to as full recall and the other which is heavily weighted toward recent prices seen as a reference price effect after its use in the marketing literature).

In the remainder of this section we discuss each of the above effects in turn. We draw upon prior literature to identify the signature pattern which defines an effect and then relate this to the elements of our structural model to formalize a test for the effect's presence.

3.1 Position-dependent effects

Preference learning

The standard SP assumption is that respondents know their preferences regarding the goods on offer prior to making their first valuation response (i.e. prior knowledge and within-survey information is sufficient to generate preferences which remain stable throughout the valuation responses). However, Plott's (1986) Discovered Preference Hypothesis (DPH) emphasizes the dynamic process of preference learning through practice and repetition. Arguably, the preference discovery process may in part occur during the process of repeated choice making if the act of assessing trade-offs focuses respondents' attention onto the various attributes of the good. Both DeSarbo, et al. (2004) and Bateman, et al. (2008a) provide suggestive evidence of preference discovery effects within SP studies², although both focus on goods which are somewhat novel and hence more likely to be subject to preference discovery than more familiar goods.

There are two results which might indicate preference learning within a repeated SP experiment. First, given that we can say with certainty that respondents already know their preferences for money prior to the survey a signature pattern for preference discovery might be stability in the utility of income accompanied by changes in the utility of the non-money attributes as respondents discover and alter their preferences regarding a good as they progress through a series of SP questions (i.e. there will be significant order effects in the $\beta(j)_i$ for those non-money attributes). Second, if preference learning merely confirms and strengthens initial attitudes towards a good then an additional indicator might be to test for significant reductions in the variance of the residual term $e(j)_{\omega,ij}$. Of course it is feasible that both patterns might occur as an SP question order progresses if respondents both alter and refine their preferences for a good during the response process. Given this we conduct both tests, investigating potential changes in the utility of the non-price attributes and patterns in the variance of the residual term as the question order progresses.

Institutional learning

² Arguably the findings of Brown, et al. (2008) and Amir and Levav (2008) might also be reinterpreted within a DPH framework.

Swait and Adamowicz (2001) argue that the first choice tasks in a sequence of DCE questions may provide a way to learn about the task format. The standard formulation here is also a reduction in the overall error variance similar to the more general form of preference learning. However, such learning is seen as distinct from the preference formation process envisaged by Plott and conforms to the idea of ‘institutional learning’ described by Braga and Starmer (2005). In a recent examination of repeated question CV formats, Bateman et al., (2008a) explicitly test for institutional learning finding it to be a very fast process which is completed in the transition between the first and second valuation question. In effect, when faced with the first task respondents may not anticipate a second question. The latter study suggests that they adjust to this format once the second task is presented and effects associated with institutional learning become insignificant at that point. Whether or not such effects will always be so discrete may well be an open empirical question, but certainly a signature pattern for institutional learning would be changes in the parameters of the model that are focused on the early tasks of a question sequence.

Fatigue

Swait and Adamowicz (2001) postulate that later tasks within a DCE question sequence might induce boredom or fatigue, a problem which provides a major focus of the extant literature on ordering effects. This issue has most generally been investigated through considerations of patterns within the error variance (sometimes referred to as the ‘scale parameter’) with increases as an ordering progresses being interpreted as evidence of fatigue. However, a number of studies have failed to find evidence of significant fatigue effects including Savage and Waldman’s (2008) comparison of mail versus online surveys, a study by Ohler, et al. (2000) who examine responses to 80 binary choice tasks finding no evidence of fatigue and Brazell and Louviere (1997) who find no increase in error variances until they exceed 64 choice sets.

This of course does not mean that fatigue should be ignored and several commentators argue that it is interaction of choice and design complexity with the cognitive abilities of the respondent to perform the task which will determine the presence and degree of fatigue affects (Scarpa, et al., 2003; Dellaert, Brazell and Louviere, 1999; DeShazo and Fermo 2002; Caussade, et al., 2005; Hensher and Rose, 2009). However, given that, for reasons explained previously, we have simple choice tasks involving choosing between a constant option and one varying alternative option, this should ensure a low cognitive burden throughout the exercise thus mitigating fatigue effects. This is further mitigated by using a number of choice tasks well

below the threshold for adverse impacts suggested by Brazell and Louviere. However, to confirm this we test for fatigue by examining order related patterns in the variance of the residual term, $e(j)_{\omega,ij}$, across the question ordering

3.1 Precedent-dependent effects

Starting point effects

Within the SP literature, starting point effects³ were first observed within iterative bidding games where it was observed that final valuations were biased towards the randomly assigned initial bid amounts in open-ended questions (Rowe, et al, 1980; Boyle, et al., 1985; Mitchell and Carson, 1989). A variant of this effect has been found in repeated testing of the double bound dichotomous choice (DBDC) format (Hanemann, et al., 1991) with numerous studies finding a clear effect of starting point upon subsequent responses (Holmes and Kramer, 1995; Herriges and Shogren, 1996; Bateman, et al., 2001; O’Conor, et al., 1999; Whitehead, 2002; Chien, et al., 2005; Alevy, et al., 2008). A number of choice experiment studies have used split sample approaches to investigate starting point effects. Evidence of such effects are reported by Carlsson and Martinsson (2008), Holmes and Boyle (2005) and Ladenburg and Olsen (2008) who find that samples that are shown lower prices in the initial question of an ordering yield significantly lower WTP values.

One possible explanation for observations of starting point bias is that respondents come to an SP exercise with malleable preferences. As a result, the attributes of the options in the opening task act as an informational cue that is highly influential in the construction or shaping of those preferences. Respondents’ choices in subsequent tasks are framed or anchored by a comparison of the attributes of the option under consideration and the attributes of the options presented in the opening task. Accordingly, the starting point hypothesis defines a clear signature pattern wherein respondents will tend to regard options encountered in subsequent tasks more (less) favorably if they appear to be better (worse) than the options offered in the initial task. Formalizing this within the framework of our structural model, starting point effects can be described as a calculation of a precedent variable $P(\mathbf{X}_{\omega}; \mathbf{X}_{A_1})$ that captures a straight comparison of the current option with the options presented in the initial task. A difficulty with the anchoring hypothesis is that it may be hard to distinguish from strategic behavior (Carson and Groves, 2007)⁴, such as the respondent treating the initial price seen as the most likely

³ These are often seen as a form of the more general anchoring phenomena (Tversky and Kahneman, 1974).

⁴ To see this, note that pure anchoring occurs when an individual conditions an action on a stimulus, such as a randomly generated number, even though the individual knows the stimulus has no informational content with relevant to the decision being made. If the individual believes the stimulus contains information relevant to the

actual cost and adjusting subsequent responses relative to this⁵. Such confounding makes definite interpretation of starting point effects challenging⁶.

Strategic behavior

Carson and Groves (2007) show that an SBC question is incentive compatible and hence proof against strategic behavior provided that the characteristics of the WTP scenario credibly entail: (a) a take-it-or-leave-it proposition on obtaining the good in question at a single specified price, (b) a consequential response in the sense that respondents perceive that the government will take the survey results into account in its decision making, and (c) that the government has the ability to compel payment at the stated price if the good is provided. Certain of these conditions become compromised within designs which confront the subject with multiple choice tasks (thereby clearly contravening condition (a) and arguably undermining condition (c)). The simplest theoretical prediction in the pure public goods case is that the single and repeated choice formats need not produce different estimates⁷. More specifically, respondents seeing different prices for the same (or close to the same) good, or seeing the same price for different goods, may attempt to exploit this flexible pricing to increase consumer surplus.⁸

While it could be argued that strategic behavior might not be a problem in repeated choice experiments due to the complexity of the strategic task demanded of the individual, Carson and Groves (2007) point out that there may be simple strategies that respondents can use that are optimal or close to optimal. These typically revolve around rejecting a preferred alternative in a given choice set when a better deal was available in a previous choice set. This

decision, then it may be optimal to condition that decision upon that stimulus. The presence of a statistical relationship between the stimulus and the decision made can thus be consistent with either pure anchoring or standard economic behavior depending on the belief structure of the individual with respect to the informational content of the stimulus.

⁵ E.g. by perceiving the actual cost to be some weighted average of the initial cost and that stated in a given option.

⁶ This is particularly true in the DBDC format (DeShazo, 2002; Whitehead, 2002) since only one subsequent response can be influenced by the initially offered price. Carson and Groves (2007) argue that the main neoclassical prediction related to asking the second DBDC question is that the perfect correlation between the WTP distributions implied by responses to the first and second questions should be broken due to changes in either the implied information set or the change in the strategy space as long as the two questions are linked. More structure has to be placed on the problem to obtain a sharper prediction.

⁷ Respondents in the repeated choice case need to make some type of inference about how responses over multiple questions will be combined to help determine the single level of the public good to be provided so that truthful preference revelation may no longer be an optimal strategy for all respondents (Carson and Groves, 2007; Taylor, Morrison and Boyle, 2007). Since the key pure public good property, that all agents share the same level of the good, is fundamental to the specific nature of the incentives respondents face, it is difficult to infer anything about the incentive properties of multiple question formats when used with pure public goods from those using quasi-public or private goods where the incentive structure can be quite different. Using such formats for public goods provided by voluntary payments may further disturb behavior due to additional changes to the incentive structure induced by the payment condition.

⁸ A further possibility, only partly explored here (through consideration of variability), is that respondents react to multiple choices by questioning the credibility of the exercise.

strategy provides the signature pattern required to identify such behavior. Specifically strategic responses will reveal a pattern wherein respondents are more likely to reject a given option when the deal it offers is not as good as one previously seen.

As one moves to a reasonably long sequence of questions, it is possible to define a number of forms of strategic behavior. These include the strategic interpretation of the starting point effect mentioned above. However, a more conventional depiction of strategic behavior arises from models of price dispersion which sought to explain how the same good could sell for different prices at stores situated close to each other. Stigler (1961) put forth the core idea that at least some consumers might strategically search for price information with recall (see also Rothschild, 1973)⁹, a key feature of this process being that consumers were assumed to have full recall of all prices previously seen (Stahl, 1989). To capture such “*full recall*” strategic behavior we can define a precedent variable with the general form, $P(\mathbf{X}_o; \mathbf{X}_q)$; that is to say a variable that compares the attributes of the current option, \mathbf{X}_o , to those of the option that the respondent currently believes to have offered them best deal, \mathbf{X}_q .

Reference price behavior

While micro theorists worked with consumers with perfect recall, early statistical work had already shown (Gabor and Granger, 1961) that the real world consumer was more fallible¹⁰. Marketing researchers took note and developed an extensive literature on the effects of deviations between an offered price and consumer’ perceptions of the ‘reference price’ of a good. Typically consumers form reference prices from observed prior prices through a process which was initially modeled using standard econometric formations for adaptive and rational expectations (Winer, 1985). Later work has tended to emphasize reference price formation as a memory based recall process (Briesch, et al. 1997; Mazumdar, et al. 2005) with more recently observed prices playing the primary role.¹¹ The impact of this reference price was first captured though Wicksteed’s (1910) intuition that buyers’ behavior is significantly affected by whether

⁹ Salop and Stiglitz (1977) worked out the equilibrium spatial distribution of prices with a mixture of informed and uninformed consumers. Varian (1980) developed a theory of why stores had occasional big sales rather than making a series of small price changes with a similar mixture of consumers to explain the observed pattern in intertemporal variation in prices without invoking a change in cost or overall consumer demand.

¹⁰ The emphasis in the economics literature has been on heterogeneity in terms of either the gain from search or with respect to search cost. Stahl notes that adding a cost of reverting to previously offered prices would alter the equilibrium nature of price dispersion but does not pursue this issue. A recent unpublished paper by Janssen and Parakhonyak (2008) shows that adding a cost of such reversion substantially alters the key results on consumer search in the sense that there is no longer a single reservation price even if the consumer knows the parameters of price distribution. This cost of returning to previous offers in some ways mimics the less than perfect recall we consider in the next section.

¹¹ Simple variants of these processes look like adaptive expectations models where options are down weighted the farther in the past they occur and are forgotten relatively quickly.

they regard offers as being ‘good deals’ or ‘bad deals’ relative to previous options.¹² This stands in contrast to the standard assumption made in a sequential SP exercise that any choice is assessed independently of preceding options. Evidence for significant reference price effects is so pervasive that they are accepted as an empirical generalization within this literature (Kalyanaram and Winer, 1995).¹³

Reference pricing effects should present a distinctive signature pattern. They should manifest as a relationship between an option given in a particular choice and a ‘reference deal’ reflecting preceding options. Specifically we would expect to observe options offering deals that are better (worse) than the reference deal being regarded more (less) favorably than can be explained simply as a consequence of the attributes of that option. In the context of our structural model, a reference deal constructed from the last m observed tasks can be described by a precedent variable of the form $P(X_{\oplus}; X_{A_{j-1}}, \dots, X_{A_{j-m}})$ and we discuss this specification subsequently. Note that one can view reference pricing within a strategic lens and that when m includes all of the offers seen and they are equally weighted then reference pricing is equivalent to full recall strategic behavior.¹⁴

4. Design and Survey Implementation

This is a primarily methodological study and all elements of design were intended to assist the production of clear and interpretable results. This included the decision to use the quality of domestic tap water supplies as the public good to be valued. Tap water quality attributes are ideal for this exercise for a variety of reasons. First, both surveys of customer opinions (including in the survey area of Norfolk, UK) and objective assessments show that the issues defining variation in the quality of supply are not complex or emotive problems such as health risks¹⁵ but rather the more readily understood issues of water discoloration due to soils

¹² The difference between the reference price and actual price has been labeled as transactional utility by Thaler (1985) and for a positive difference can be thought of as the gain in consumer surplus from buying now rather than later. As one might expect consumers develop a “promotion expectation” with respect to these gains that can influence their purchase decisions (Bucklin and Lattin, 1989). From a prospect theory perspective (Tversky and Kahneman, 1991) the really bad situation for a firm is for the current price to be above the consumer’s perceived reference price for the good so that the consumer takes a “loss” from buying now. A formal theory of reference price behavior from an economic perspective is developed by Putler (1992) which allows for the possibility of asymmetric effects due to loss aversion. The empirical literature has been mixed on whether the empirical finding of a dependence of current behavior on past prices is driven by some type reference price effect or by stockpiling behavior (Slonim and Garbarino, 2009).

¹³ Perhaps one of the clearest indicators of this issue is the common practice of displaying of “regular” price next to “sale price” or having a percent-off sale (Mayhew and Winer, 1992).

¹⁴ While the focus here is on the typical variant of reference pricing which tends to place greater weight upon recently observed costs, it possible that the reference price in a sequential SP exercise is set by the cost amount in the first choice set in which case it is indistinguishable from a starting point effect.

¹⁵ Tap water in the United Kingdom is considered amongst the best in the world in terms of consistency of supply

and odor/taste problems typically due to excess chlorination (MORI, 2002; Hunter, 2003). Second, those same surveys suggest that many if not most respondents are likely to have noticed variation in these attributes previously. Third, climate change and increasing population pressures are likely to increase pressure upon these aspects of water supplies in the future (Kabat, *et al.*, 2002; Holman, *et al.*, 2002)¹⁶, generating scenarios of provision change for valuation purposes. Finally, for some years now UK water consumers have paid for maintenance and improvements to services through increases to water bills, a coercive payment vehicle.

Given this background, we developed scenarios that varied water discoloration and odor/taste attributes of water quality alongside changes in water bills. We used focus groups to refine the scenarios to be used in the survey. These suggested that water supply states can be described by three attributes: (1) the number of days annually where a household's tap water smelled and tasted of chlorine (ODOR)¹⁷, (2) the number of days annually where the tap water was a rusty color (COLOR) and (3) the addition to the household's annual water bill induced by implementing technical procedures to address these problems (COST).¹⁸ Combinations of different levels of each of these attributes make up our SP options.

The decision to use tap water quality, described by three easy to understand attributes in a binary choice against the status quo, in conjunction with the decision to use high quality personal interviews may mitigate against the observation of some ordering effects, such as fatigue.¹⁹ However, given the methodological focus of our study, this is not seen as a problem. Less familiar goods and/or more complex choice situations without the motivation of an interviewer being present may indeed yield empirically different ordering effects, but it is the development of a methodology to detect order effects which is a key intended contribution of this paper. To that end, the choice of the domestic water supply good reduces non-focal complexities. One of these is that the non-cost attributes of the chosen good share common units; both ODOR and COLOR being measured in the number of days per year that these problems arise. This has considerable advantages when testing for precedence-dependent

and health risk levels (Hunter, 2003). This is due to a combination of naturally occurring factors (e.g., rainfall and temperature) as well as high investment and monitoring levels.

¹⁶ Climate induced changes in weather patterns (especially higher peak rainfall) are resulting in increased erosion and loadings of silt into waterways leading to discoloration problems. Higher rainfall also increases diffuse pollution loadings and fecal matter runoff from farms as well as more incidents of sewage treatment work overflows (exacerbated by increasing populations in the sampled region). These necessitate greater levels of chlorination leading to more instances of odor/taste problems. We are grateful to Irene Lorenzoni (UEA, UK) for conversations regarding public perceptions of climate change issues prior to our focus group work.

¹⁷ Incidents of smell and taste problems are too collinear to be separated, so were treated as a single attribute.

¹⁸ The household water bill has highly desirable properties as a payment vehicle (Mitchell and Carson, 1989), being universal and non-voluntary.

¹⁹ The use of such simple options should help avoid potential problems that have been identified when respondents face options with a large number of attributes (Maddala, Phillips and Johnson, 2003).

anomalies. As the typical number of days per annum of each attribute is low it seems reasonable to assume (and trivial to test for) that per-day unit costs are relatively simple for respondents to estimate (albeit undoubtedly with some error) and hence deals readily assessed. We can then examine the impact of preceding deals as set out in the previous section.

Given our methodological focus we also adopt the most straightforward choice format. This involves a choice between a constant status quo (SQ) option in which no intervention is undertaken and costs remain unchanged, and a single ‘alternative’ state, varied across choice questions²⁰. The SQ option was defined as the likely level of tap water problems to be experienced over the coming year. Focus group tests indicated that credible levels were 10 days per year for the ODOR attribute and 5 days per year for the COLOR attribute in the absence of any intervention (a zero increase in the water bill) to address these problems. Following discussion with water quality scientists²¹ regarding the feasibility of preventative works, we assigned four levels to each attribute: ODOR (0, 3, 6 and 10 days); COLOR (0, 1, 3, and 5 days), and COST (£10, 20, 30 and 50). Figure 1 illustrates a typical choice question where the SQ option is shown under the ‘No Scheme’ heading and an example alternative option is given under the heading of ‘Scheme A’²².

INSERT FIGURE 1 ABOUT HERE

One aspect of our study design arose from a desire to supplement conventional parametric analysis with a more robust non-parametric approach to testing for the presence of ordering effects. To achieve this we first define two options, one of which has relatively larger improvements in both the ODOR and COLOR attributes, while the other concerns a smaller improvement (these are subsequently referred to as the ‘larger’ and ‘smaller’ goods respectively). Combining these with all four levels of the COST attribute defines eight distinct options. We then ensure that for every respondent their first choice question involved the choice between the ‘no scheme’ option and a random draw of one of these eight alternative options. This in itself provides a simple test of sensitivity to changes in price and quality attributes which can be conducted upon just the first response of each subject. However, the main purpose of this approach is realized by, for each respondent, ensuring that the first question they are presented with is then repeated as their final choice task. Comparison of this first and last response to the

²⁰ The use here of a sequence of stated paired comparisons against the SQ allows us to isolate the independence issue from other potential strategy issues that are involved with the use of multinomial choice questions with pure public goods.

²¹ We are particularly grateful to Professor Glen George, Centre for Ecology and Hydrology, Lancaster, UK.

²² As is standard in such exercises, respondents were instructed that each question should be considered independently of any preceding questions. The precise instruction used was as follows: “Please consider each choice on its own merits, irrespective of whether previous questions gave better or worse deals. Treat each choice as the only choice available to you”.

same question provides a highly robust, within-person, non-parametric test for ordering effects.²³

A further important element of our study was to ensure that it avoided the possibility of design induced ordering effects. Bateman et al. (2008b) detail a previously unrecognized problem that in testing for order related issues uniquely tied to error variance such as fatigue or learning that the mix of choice sets seen at each order needs to be identical. That is exactly true by construction in this study for the first and last questions present the respondent with the same options and hence share the same covariance matrix elements. For the intermediate questions efforts were made to generate a balanced mix of choice sets across the question order by using a Latin square design (Street and Street, 1987) to rotate options across the sample. This takes a given question sequence (say a,b,c,,,x,y,z) presented to Respondent 1 and moves the first question to the last (to yield b,c,d,,,y,z,a) to provide the sequence for Respondent 2²⁴. Iterating this approach repeatedly through the sample generates a much more even mix of options across the question order than typically seen in SP studies.

A further potential problem in the form of misspecification bias was addressed by employing a full factorial design²⁵ to determine all possible choice tasks for our attributes and their levels. Although full factorial designs are relatively rare in environmental applications they allow the analyst to estimate all main and interaction effects independently of one another.

Prior to the main survey a series of pilot surveys and one-to-one debriefing exercises helped refine the wording and presentation of the questionnaire such that tasks were readily understood by survey respondents and not unduly demanding. A team of experienced interviewers that were specially trained in how to administer the survey conducted face-to-face interviews at respondents' homes. Respondents' addresses were selected on a randomized basis across the study area (East Anglia, UK) so as to ensure variation across household socio-

²³ Following Bateman, et al., (2004) respondents were further randomly assigned in equal proportions to one of two treatments: the ADV treatment where all of the price and non-money attribute levels respondents will encounter were shown in advance; and the stepwise (STP) treatment where attribute levels are only revealed as respondent progress through the questionnaire. Most sequential SP studies generally fall between these two extremes with the STP treatment being perhaps the mostly used practice. Space constraints prevent a separate analysis of these two subsamples so they are combined for the purposes of this paper.

²⁴ To further control for possible design induced ordering effects the possible options were allocated into two blocks. In the first block options involving the 'extreme' (i.e. highest and lowest) levels of the attributes filled the first half of each respondents' choice tasks with the remaining 'non-extreme' options allocated to the second half of tasks. In both blocks Latin squares were used to rotate tasks. For the second half of the sample the 'non-extreme' Latin squares preceded the 'extreme' Latin square. We feel that this approach provides a much more balanced mix of choice sets across orders than typically seen but there were still some potentially confounding divergences across particular orders. These idiosyncrasies are avoided here by combining orders. This potential problem can be avoided in future studies by implementing another layer of design to ensure that the same mix of choice sets appears at each order.

²⁵ See Louviere, Hensher and Swait (2000), for further discussion of full factorial designs. The full factorial approach avoids many common design problems but is only practical for experiments with a small number of attributes and levels. For full details of the design used in the present study see Bateman et al., (2008b).

economic and demographic characteristics. The different versions of the survey were randomly assigned to respondents and a sample of 864 completed questionnaires collected.

5. Econometric Model

As noted previously, our experiment presents respondents with a series of binary choice tasks in which they are asked to choose between the *status quo* (SQ) level of tap water problems and some *alternative* that comes at a price.

We indicate the SQ by the subscript 0 and specify the utility it offers as;

$$U_{0,ij} = \mathbf{X}'_0 \boldsymbol{\beta}(j) + \mathbf{Z}'_0 \boldsymbol{\gamma} + \mathbf{P} \left(\mathbf{X}_0; \mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1} \right)' \boldsymbol{\theta} + e(j)_{0,ij} \quad (5)$$

As per our specification of the utility function for a generic choice task option (4), \mathbf{X}_0 in (5) is the vector describing the attribute levels; in this case the status quo levels of the two water quality attributes and an associated price which is always zero. Notice, however, that (5) differs from (4) in a number of ways. First, for simplicity (5) assumes that the preference parameters, $\boldsymbol{\beta}$, and the precedent variables parameters, $\boldsymbol{\theta}$, are constant across individuals. In addition, (5) is not restricted to the examination of the main effects of the treatment variables captured by the $\boldsymbol{\beta}$ parameters. Rather our use of a full factorial design allows us to capture possible interaction effects through the inclusion of the cross-products of the attribute levels in the vector \mathbf{Z}_0 . The parameters on those interaction terms, $\boldsymbol{\gamma}$ are assumed constant across respondents and, for simplicity, we ignore the possibility that these parameters may also exhibit position-dependent order effects. Finally, for the purposes of econometric identification, the location of the utility scale is set through the normalization $\alpha_{0,i} = 0$ ($i = 1, 2, \dots, N$).

The utility from the alternative option is:

$$U_{\omega,ij} = \alpha(j)_i + \mathbf{X}'_{\omega} \boldsymbol{\beta}(j) + \mathbf{Z}'_{\omega} \boldsymbol{\gamma} + \mathbf{P} \left(\mathbf{X}_{\omega}; \mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1} \right)' \boldsymbol{\theta} + e(j)_{\omega,ij} \quad (6)$$

Given the fact that each of our alternative options can be completely described by the attribute levels contained in \mathbf{X}_{ω} , we make the simplifying assumption that $\alpha(j)_{\omega,i} = \alpha(j)_i$ ($\forall \omega \in \mathbb{A}; \omega \neq 0$). That is to say, we assume that the fixed level of utility from choosing the alternative option does not depend on the particular option being offered. A positive value for $\alpha(j)_i$ would signal a general preference for the change offered by choosing the alternative, a negative value a preference for the continuity offered by maintaining the status quo.

The difference in the utility offered by the two options is then;

$$\begin{aligned}
u_{ij} &= U_{\circ,ij} - U_{0,ij} \\
&= \alpha(j)_i + \mathbf{x}'_{\circ} \boldsymbol{\beta}(j) + \mathbf{z}'_{\circ} \boldsymbol{\gamma} + \mathbf{P}_j \left(\mathbf{X}_{\mathbb{A}_j}, \mathbf{X}_{\mathbb{A}_{j-1}}, \dots, \mathbf{X}_{\mathbb{A}_1} \right)' \boldsymbol{\theta} + \varepsilon(j)_{ij}
\end{aligned} \tag{7}$$

where

$$\mathbf{x}_{\circ} = \mathbf{X}_{\circ} - \mathbf{X}_0$$

$$\mathbf{z}_{\circ} = \mathbf{Z}_{\circ} - \mathbf{Z}_0,$$

$$\mathbf{P}_j = \mathbf{P} \left(\mathbf{X}_{\circ}; \mathbf{X}_{\mathbb{A}_j}, \mathbf{X}_{\mathbb{A}_{j-1}}, \dots, \mathbf{X}_{\mathbb{A}_1} \right) - \mathbf{P} \left(\mathbf{X}_0; \mathbf{X}_{\mathbb{A}_j}, \mathbf{X}_{\mathbb{A}_{j-1}}, \dots, \mathbf{X}_{\mathbb{A}_1} \right), \text{ and}$$

$$\varepsilon(j)_{ij} = e(j)_{\circ,ij} - e(j)_{0,ij}.$$

Note that the specification in (7) allows for heterogeneity in preferences through the person-specific fixed effects $\alpha(j)_i$ and through the error term, $\varepsilon(j)_{ij}$.

Position-Dependent Ordering Effects

Position-dependent order effects enter (7) in both the fixed effect, $\alpha(j)_i$, and the preference parameters $\boldsymbol{\beta}(j)$, as well as through the residual term $\varepsilon(j)_{ij}$. To introduce position-dependence into the fixed effect and the preference parameters, we construct trend variables which allow the value of those parameters to be a function of position in the series of tasks. Rather than imposing a linear trend, we introduce a further parameter λ , which provides a power transformation of j of the form j^λ . The shape of the trend is now determined by the value of λ . For example, if $\lambda = 1$ then the trend is linear, if $\lambda = -1$ then the trend takes the shape of an inverse function, while if $\lambda = 0$ there is no trend,

The fixed effect is then specified as;

$$\alpha(j)_i = \alpha_i + \alpha + j^{\lambda_a} a \tag{8}$$

and the preference parameters as;

$$\boldsymbol{\beta}(j) = \boldsymbol{\beta} + j^{\lambda_b} \mathbf{b} \tag{9}$$

Observe that by estimating different shape parameters, λ_a , and λ_b , our model permits the fixed effects to exhibit a different form of position-dependent order effect to the preference parameters.

According to the preference learning hypothesis we might expect to see changes in the preference parameters as we progress through the series of tasks. The specification in (8) and (9) would capture such change as some non-zero value of λ coupled with one or more significant parameters amongst \mathbf{b} and a .

To introduce position dependence into the error terms, we allow for the fact that the variance of those terms may change over the course of the SP exercise. We capture such changes through the estimation of a parameter, σ_j , giving;

$$\varepsilon(j)_{ij} = \varepsilon(\sigma_j)_{ij} \quad (10)$$

Rather than imposing a trend on the scale parameter, we allow separate scale parameters to characterize the error variance at different positions in the sequence. As described in the previous section, a preference learning hypothesis might assume greater acuity in the expression of preferences as respondents progress through the series of choice tasks. Such a pattern of behavior would express itself as a declining trend in the scale parameters. In contrast, a fatigue hypothesis reasons that respondents lose concentration as they progress through the tasks. That pattern would express itself as an increasing trend in the scale parameters. A combination of the two can yield a U-shaped pattern.

Precedent-Dependent Ordering Effects

We wish to construct differenced precedent variables, $\mathbf{P}_j(\mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1})$, capable of capturing the three separate forms of precedent-dependent ordering effect identified:

- Starting point effects
- Full recall price behavior
- Reference price behavior

Since each task in our experiment is a binary exercise in which an unchanging SQ is contrasted with some alternative option we assume that the attributes of options in each choice set can be captured by the net difference between the attributes of the alternative and the SQ. Accordingly, we make the simplifying assumption that $\mathbf{P}_j(\mathbf{X}_{A_j}, \mathbf{X}_{A_{j-1}}, \dots, \mathbf{X}_{A_1}) = \mathbf{P}_j(\mathbf{x}_\emptyset, \mathbf{x}_{\mathbb{Q}_{j-1}}, \dots, \mathbf{x}_{\mathbb{Q}_1})$, where $\mathbf{x}_{\mathbb{Q}_k} = \mathbf{X}_{\mathbb{Q}_k} - \mathbf{X}_0$ and $\mathbb{Q}_k \in \mathbb{A}$ denotes the option presented as the alternative to the SQ in choice task k ., where

The vector $\mathbf{x}_{\mathbb{Q}_k}$ indicates the cost and the days of water quality improvements of the two types offered by the alternative in task k . To further simplify, we choose to summarize the quality of the options offered in task k , by calculating the cost per day of tap water improvements that it offers. Let us denote that daily cost by $p_{\mathbb{Q}_k}$ or more simply just p_k . Observe that p_k can be interpreted as the unit price of improvements offered by an option²⁶. In essence

²⁶ In the present analysis that price does not distinguish between color and odor improvements. Tests indicated that

p_k approximates the quality of the deal offered by that option. Low values for p_k represent good deals where improvements can be purchased at relatively little cost. High values for p_k represent bad deals where the price of improvements is comparatively onerous. Making that further simplification, our precedent variables take the general form, $\mathbf{P}_j(p_j, p_{j-1}, \dots, p_1)$.

Let us now consider how to construct a precedent variable to capture the possibility of strategic misrepresentation of preferences. When considering how to respond to a particular choice task, a strategic respondent must evaluate a number of factors including how likely it is that the regulator will actually provide the package offered. In that regard, one perfectly reasonable assumption would be that the regulator is prepared to provide any of the options presented in the SP exercise. Why otherwise would the respondent have been asked to express an opinion on them? It follows that the optimal response for the strategic respondent is to signal to the regulator that they will not accept an option unless it represents the best deal on offer. Accordingly, even if a respondent recognizes an option as affording a welfare gain, they may still choose to misrepresent their preferences and refuse that offer.

Of course, progressing through the series of choice tasks, a respondent's perception of what is the best deal depends on which options they have been offered in preceding tasks. Since our experimental design introduces significant variation in task order across respondents, our data allows for identical options to be perceived by some as the best deal offered so far and by others as worse than the best deal. That variation provides us with the possibility of identifying strategic behavior. We create an indicator variable that takes a value of one if the deal offered by the current task is worse than the best deal observed so far and equal to zero otherwise. We call this a worse-deal indicator variable which we define as;

$$S_j = I(p_j > \min p_{j-k}, k = j - 1, j - 2, 1) \quad (11)$$

Where $I(\cdot)$ is the indicator function which equates to 1 if the term inside the brackets is true and 0 otherwise. Strategic responses should be manifested through an increased tendency to say "No" to an offer when it is worse than the best deal observed so far.

The starting point hypothesis contends that the deal offered in the opening choice task is highly influential in the construction of respondents' preferences. As a result, respondents may respond to subsequent choices as if their value for tap water quality improvements has been shaped by the deal offered in the opening choice tasks. Accordingly, deals encountered in

this produced a unit price to which respondents were price sensitive. However, one could relax the implicit equality assumption if desired.

subsequent tasks that are better (worse) than those offered in the initial tasks will tend to be regarded favorably (unfavorably). In mathematical terms it is useful to denote Q_j as

$$Q_j = p_1 - p_j \quad (12)$$

In the case of the anchoring hypothesis, the difference $p_1 - p_j$ gives an indication of how the price of improvements in the current task compares to a respondent's WTP that has itself been shaped by the prices offered in preceding tasks.

The reference price hypothesis contends that respondents' choices are partly determined by how favorably the deal being offered in the current task compares to the amount a respondent expects to pay for a day's improvement in tap water quality.

That expectation might take a number of forms, it might simply be the deal offered in the last choice task, it might be an average of deal offered in all previous tasks or it might more heavily weight recent, or for that matter early, tasks. Accordingly we calculated the expected deal as the weighted average of deals observed in previous tasks;

$$\bar{p}_j(\eta) = \sum_{k=1}^{j-1} \frac{k^\eta}{\sum_{m=1}^{j-1} m^\eta} p_{j-k} \quad (13)$$

Observe that the weights used in calculating this average are a function of a power transformation parameter, η , that we seek to estimate from the data.

- If η takes a value of 0, then the expectation reduces to the straightforward average of previous deals observed.
- If η takes on a positive value then the expectation places greater weight on deals observed in early tasks.
- If η takes on a negative value then the precedent places greater weight on deals observed in more recent tasks.

Whatever the value of η , reference pricing behavior will manifest itself as choices that are in part guided by the difference between the expected deal and the current deal;

$$R_j = \bar{p}_j(\eta) - p_j \quad (14)$$

The difference R_j indicates the extent to which respondents frame an offer as being a good or a bad deal compared to their current reference deal; when R_j is greater than zero the current deal is good relative to the reference while if R_j is less than zero the current deal is bad

relative to the reference. Under reference price behavior, consumers will tend to prefer a deal when the nature of preceding options has R_j taking on a larger rather than smaller value

Inserting (9), (10), (11), (12) and (14) in (8) provides our final specification of the utility difference equation;

$$\begin{aligned} u_{ij} &= \alpha_i + \alpha + j^\lambda a + \mathbf{x}'_{\ominus}(\boldsymbol{\beta} + j^\lambda \mathbf{b}) + \mathbf{z}'_{\ominus} \boldsymbol{\gamma} + \theta_1 S_j + \theta_2 Q_j + \theta_3 R_j + \varepsilon(\sigma_j)_{ij} \\ &= \alpha_i + v_j + \varepsilon(\sigma_j)_{ij} \end{aligned} \quad (15)$$

Given (15), a respondent's choices will be governed by the following rule;

$$\text{if } u_{ij} \begin{cases} \leq 0 \text{ then choose } \textit{Status Quo (option 0)} & (y_{ij} = 0) \\ > 0 \text{ then choose } \textit{Alternative (option } \ominus) & (y_{ij} = 1) \end{cases} \quad (16)$$

where y_{ij} is the dependent variable indicating whether individual i on choice occasion j selected the SQ or the alternative option.

If we assume that $\varepsilon_{ij} \sim i.i.d. N(0, \sigma_j^2)$ then the choice probabilities become;

$$\begin{aligned} Pr[y_{ij} = 0] &= \Phi\left(-\frac{\alpha_i + v_j}{\sigma_j}\right) \\ &\text{and} \\ Pr[y_{ij} = 1] &= 1 - \Phi\left(-\frac{\alpha_i + v_j}{\sigma_j}\right) = \Phi\left(\frac{\alpha_i + v_j}{\sigma_j}\right) \end{aligned} \quad (17)$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal and the equality in the final line follows from the symmetry of that function.

The last step in formulating the model is to complete the specification for the random effects, α_i . Following standard practice, we assume that these are distributed normally across the population; that is to say, $\alpha_i \sim i.i.d. N(0, \sigma_\alpha^2)$. To fix the scale of the model we normalize by setting $\sigma_\alpha^2 = 1$. In that case, the likelihood contribution of individual i is;

$$l_i = Pr[\mathbf{y}_i] = \int_{-\infty}^{\infty} \phi(\alpha) \prod_{j=1}^J \Phi\left(q_{ij} \left(\frac{v_j}{\sigma_j} - \sqrt{\frac{\rho_j}{1 - \rho_j}} \frac{\alpha}{\sigma_j}\right)\right) d\alpha \quad (18)$$

where:

\mathbf{y}_i is the J -vector formed by stacking the dependent variables, y_{ij} , for individual i .

$q_{ij} = 2y_{ij} - 1$ is a transformation of the dependent variable that simplifies presentation and calculation.

$\rho_j = 1/(\sigma_j + 1)$ is the correlation between the responses of the same individual resulting from the person-specific random effect.

$\phi(\cdot)$ is the probability density function of a standard normal.

The integral in (18) can be estimated using Gaussian quadrature and the log-likelihood function to be maximized with respect to the parameters is then:

$$\ln L(\alpha, a, \boldsymbol{\beta}, \mathbf{b}, \gamma, \boldsymbol{\theta}, \eta, \lambda_a, \lambda_b, \boldsymbol{\sigma}) = \sum_{i=1}^N \ln l_i \quad (19)$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ are the vectors of precedent-dependent and position-dependent scale parameters, respectively.

6. Results

To start we examine responses to the first and final choice tasks. Recall that respondents are randomly assigned to one of eight treatment groups. Individuals in each group are confronted with the same first task but those tasks differ across treatments; they offer either a relatively small or a relatively large improvement in tap water quality at one of four possible costs (£10, £20, £30 or £50).²⁷ That same task is repeated as the seventeenth and final task.

If responses to the choice tasks presented in our experiment are tapping into preferences that reflect basic conceptions of economic rationality, then we would expect those responses to display at least two fundamental patterns; individuals should prefer more of a good than less and they should prefer to pay less for that good than more. Encouragingly, the data presented in Figure 2 provides strong evidence to confirm that respondents conform to these basic precepts of economic theory when answering both the first and last choice task.

INSERT FIGURE 2 ABOUT HERE

Figure 2 plots out a series of points identifying the proportion of respondents in each treatment group to choose the alternative option. Data for first task responses is presented in the

²⁷ Analysis confirmed comparability of these samples across all collected socio-economic and demographic criteria including gender, age, income and education. Further responses ascertained that, in line with national trends (MORI, 2002) and our focus group findings, issues of taste and odor constituted the most frequent water supply problems respondents experienced. Partly as a result of this (as well as convenience) over 90% of the sample occasionally purchased bottled water. Respondents reported average annual water bills of roughly £280.

left panel, that for last task in the right panel. To aid interpretation, the point observations are joined by line segments that separately identify the large and small improvement treatment groups and approximate the path of the WTP survivor function for the two goods.

In all but two cases those line segments are downward sloping; as anticipated by economic theory, demand tends to decline as the cost of a particular improvement increases. A simple two-tailed z -test of the equality of proportions confirms that the decline observed along seven of the twelve segments is statistically significant with greater than 90% confidence. The two cases where, contrary to expectations, the segment is ascending are amongst those where the observed difference in proportions is not significant at conventional levels suggesting that no statistical violation of the monotonicity restriction occurs.²⁸

In addition, observe from Figure 2 that the proportion selecting the alternative option at a particular cost is in all cases greater when that alternative option offers the large improvement. Again, the data supports the economic prediction that, all else equal, individuals prefer more of a good thing to less. We employ a chi-squared test to confirm that the observed differences between the survivor functions for the small and large improvements are statistically significant (for the first task the χ^2 -statistic = 26.29 giving a p -value of <0.001; for the last task the χ^2 -statistic = 37.79 giving a p -value of <0.001).

Figure 2 tells a very 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. When viewed independently, there is nothing to suggest that responses to the last task (or, for that matter, the first task) are somehow inherently less meaningful in terms of revealing economic preferences.

The full story, however, may be somewhat more complex. Observe Figure 3 which plots the same data but in such a way that responses to the first and last choice tasks can be directly compared. It is immediately evident that in all cases there is a marked downward shift in the proportion of respondents choosing the alternative option in the last task. Remember that the comparison we are now making is within treatment group; that is to say, we are looking at how one treatment group responded to a particular choice task in the first question and comparing that to how that same group responded to the identical task in the final question. The appropriate statistical comparison of those responses is provided by McNemar's test of change which

²⁸ The flattening of the survival curve at the highest price point in Figure 2 above and related uptick of the survival curve for the small good below in Figure 3 are due to data from the ADV subsample and are not features present in the STP half of the sample. Direct comparisons between these treatments for both the first and last questions do not reveal any difference at the $p < .10$ level which is why the two treatments are combined (space restrictions confine further presentation of further detail to Bateman et al., 2008b). Because there are essentially the same number of observations of both treatments the analysis here should be taken as averaging over the two conditions.

confirms that the observed shift in preferences is highly significant (χ^2 -statistic = 40.56 giving a p -value of <0.001).

Our non-parametric comparisons indicate that individuals respond to choice tasks in a way that is consistent with theoretical predictions regarding economic preferences, but that those preferences change in a statistically significant fashion between the first and last tasks, with the last task responses revealing a markedly lower demand than the first task. Our non-parametric analysis, however, provides little insight into the mechanisms that drive this observed change in preferences. Our parametric analysis sets out to address that question.

INSERT FIGURE 3 ABOUT HERE

Table 1 reports estimates from our econometric analysis. The data used to estimate the model are from 833 individuals. Each individual faced at least one task that presented them with a dominating choice. Such a task comprised a costly alternative option that offered no improvements over the status quo.²⁹ Under the usual assumptions an individual who understood the task and was fully engaged in the exercise would never choose a dominated option. We observe 3% of respondents making dominated choices, a fraction small enough to be plausibly driven by recording errors but to be safe we exclude these respondents from the analysis.³⁰

Base Preference Parameters

Consider first estimates of the base preference parameters. The negative and highly significant constant, α , indicates a predilection amongst respondents to choose the status quo option, no matter what is offered by the alternative option. This is consistent either with some fraction of the sample not having a positive WTP for tap water quality improvements (or not believing they will receive them) or with what is popularly referred to as a status quo effect or bias.

The next three parameters provide estimates of the *main effects* of the treatment variables in the experiment. They can be interpreted as indicating the independent impact on utility of a marginal change in an option attribute. As expected all three are negatively signed; more days of tap water problems and higher water bills are both regarded as a bad thing.

Most analyses of discrete choice SP data, perhaps restricted by the adoption of an orthogonal experimental design, limit themselves to the estimation of those main effects. In

²⁹ In such cases the deal being offered by the task is infinitely bad. To handle this issue we need to make a reasonable assumption regarding how such deals are incorporated by respondents into their calculation of the precedent deal. We assume that the deal offered by a dominated task is considered as 20% worse than the worst non-dominated option. Plausible alternative assumptions have no impact on our qualitative conclusions.

³⁰ Results, available from the authors, using all responses are qualitatively identical to those described here.

contrast, our full-factorial design allows us investigate the possibility of interaction effects in respondents' preferences. The next three parameters in Table 1 report those interaction effects. Observe that the parameters on all three cross-product terms are statistically significant. In this case, the model would have suffered from mis-specification bias if the possibility of interaction effects had been ignored.

The interaction effect resulting from the cross-product of the two tap water problem attributes is negative, implying, quite plausibly, that the disutility of a day of one problem is magnified when coupled with inflated levels of the other problem. For example, at an option cost of £10, the marginal disutility of odor problems is -0.113 when associated with no days of color problems but -0.147 when associated with 5 days of color problems. Likewise, at the same option cost, the marginal disutility of color problems is -0.250 when associated with no days of odor problems but -0.319 when associated with 10 days of odor problems.

In contrast, the interaction effects resulting from the cross-product of each tap water problem with cost are both positive. It appears that the marginal disutility of tap water problems declines as the overall cost of the package offered by an option increases. So, for example, when enduring 5 days of color problems the disutility of an extra day of odor problems is -0.147 when the package cost is £10, -0.140 when the package cost is £20, -0.132 when the package cost is £30 and -0.117 when the package cost is £50. That result, is consistent with an increasingly large proportion of the sample encountering constraints in the appropriate budget as the package cost increases.³¹

INSERT TABLE 1 ABOUT HERE

Of course, the existence of significant interaction effects indicate that estimates of marginal WTP for tap water quality improvements will be dependent on the absolute levels of those problems and the package cost. For example, consider a state of the world characterized by 3 days of odor problems and 3 days of color problems with an associated £10 increase in water bills over the status quo. From that position, marginal WTP for one less day of odor problems is then £2.33 and for one less day of color problems is £7.88. If, however, things were somewhat worse and the number of days of color problems was 5 (instead of 3) then marginal

³¹ Consider a respondent who regards an option offering moderate levels of tap water improvements at a moderate cost as being preferable to the status quo. Imagine another option that scales-up the levels of tap 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 status quo. 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 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. In our specification, that behavior would manifest itself as an apparent reduction in the marginal disutility of tap water problems as total option cost increases.

WTP for one less day of odor problems would increase to £2.77 (from £2.33). Alternatively, if the number of days of odor problems was 6 (instead of 3) then marginal WTP for one less day of color problems would increase to £9.08 (from £7.88). Those increasing marginal WTPs reflect the negative interaction effect on the cross-product of the two tap water quality variables. The impact of the positive interaction effects on the cross-products of cost with the two tap water quality variables can be shown by considering how marginal WTP changes if, from the same baseline of 3 days of color and 3 days of odor problems, we increase the cost from £10 to £50. In that case, marginal WTP for one less day of odor problems falls from £2.33 to £1.80 and marginal WTP for one less day of color problems falls from £7.88 to £5.47. As mentioned, those falls in marginal WTP may simply reflect the fact that more respondents are inhibited by binding budget constraints at higher levels of package cost.

Position-Dependent Ordering Effects

The next set of parameters in Table 1 are those included to test for the presence of position-dependent order effects; that is to say, changes in preferences that develop over the course of a series of SP choice tasks purely as a result of position in the sequence of tasks. We capture that development of preferences through the interaction of both the constant (fixed effect) and attribute levels with trend variables (see equations (8) and (9)). The trend variables allow the development of preferences over choice tasks to take the form of any of a large class of increasing or decreasing monotonic functions.

Observe first that no statistically significant position-dependent order effects are observable in the attribute preference parameters. It appears that the disutilities derived from paying out money, suffering a day of tap water odor problems or of tap water color problems do not change relative to each other as respondents progress through the choice tasks. These observations would seem to run contrary to a preference learning hypothesis; in our application at least, respondents' preferences for attributes do not appear to change systematically simply as a result of the experience of repeatedly making choices requiring trade-offs in those attributes.

In contrast, the significant coefficients associated with the trend-interacted constant suggest that respondents do exhibit position-dependent order effects of some description. To fully understand the nature of that effect, we must first determine the shape of the trend variable. The trend in the variable is constructed as j^{λ_a} where j indexes position in the order and the shape parameter, λ_a , is estimated from the data as -1.907. As such, the trend variable declines rapidly, at a rate approximately equal to the inverse square of position in the sequence of tasks.

Recall that one interpretation of the constant in the model is as indicating a status quo effect. The development of that status quo effect at each position in the sequence of tasks can be calculated as $\alpha + j^\lambda a$ (though, through the action of α_i , respondents' individual values vary around this as a standard normal variate) and is plotted out in Figure 4. Notice that in response to the first task, preferences show little evidence of a status quo effect. However, as respondents progress through the sequence of tasks their preferences modify markedly exhibiting an increasing tendency to favor the status quo over the alternative option. Notice also though that the inverse square trend ensures that the position-dependent change in the status quo effect is heavily weighted towards early tasks; preferences change very quickly in the first few tasks and soon settle on a relatively constant status quo effect.

The particular form of position-dependent change depicted in Figure 4 is interesting. The minimal status quo effect evident in first task responses indicates that those responses attune much more closely with standard economic conceptions of preferences. Moreover, the rapid evolution of the status quo effect suggests that it is something particular about the presentation of the second and subsequent few tasks that triggers the development of a significant status quo effect.

The rapid change in the status-quo effect over the first few tasks is highly reminiscent of the patterns of change in preferences that have previously been attributed to institutional learning (Bateman, 2008a). However, explanations for the observed pattern might also be formulated from standard theory. For example, Carson and Groves' (2007) argue that increasing cost uncertainty for a public good with a coercive payment mechanism is equivalent to increasing income uncertainty and hence should reduce the propensity to pay any particular amount. The presentation of a second task may alert respondents to the possibility of cost uncertainty and that uncertainty may trigger the observed development of a status quo effect. Alternatively, the presentation of a second task might signal to respondents that the answer to the first question is not the last word. Respondents may, as a result, adjust their response strategies in subsequent tasks apparently in a manner that favors the status quo over any potential change. That behavior is reminiscent of the phenomenon modeled and explored by Corrigan, Kling and Zhao (2008). They observe that individuals are more likely to select an option if faced by a take-it-or-leave-it choice than if they are afforded the opportunity to delay their decision. Unfortunately, our experimental design does not allow us to be more precise in our suppositions regarding the cause of this position-dependent order effect.

INSERT FIGURE 4 ABOUT HERE

A second set of parameters included to examine position-dependent order effects are those which allow the scale of the error distribution to differ as respondents progress through the set of tasks. As we have seen, previous authors have interpreted an increasing scale in the error distribution as indicative of falling respondent concentration.³²

The major difficulty with the interpretation of position-dependent order effects in the scale parameter is that those effects may partially reflect differences in the mix of choice tasks presented at each position in the sequence. To see this, consider the case where a simple linear utility function is fitted to the data when the true utility function has some curvature. In this case, the error term contains both an element reflecting respondents' inability to unerringly anticipate the magnitude of the benefit from a particular option (Thurstone, 1927; Luce, 1994) and an element reflecting misspecification error. If the mix of choice sets presented to respondents at a particular position in the order contains a relatively limited range of attribute levels, then options in that task will tend to be focused in a relatively small area of the attribute space. Within that limited space, misspecification bias will itself be relatively similar such that that component of the error will exhibit substantially different variability than if the mix of choice tasks for that position contained a broader range of attribute levels.

Our experimental design was devised specifically to avoid such confounding. From questions 2 to 16, tasks were assigned to each position in a balanced manner using the Latin square design discussed previously such that the variance of attribute levels for each task should on average be equal across positions. Of course, the repeated first and final tasks, being drawn from a set of tasks limited to only one of two levels of tap water quality attributes, exhibit a greatly reduced variance. To allow for the confounding effect of those design differences, we estimate separate parameters to capture the error scales in the first and last task. We test for the possibility of position-dependent changes in scale through the middle 15 tasks by estimating three further scale parameters; one for tasks 2 to 6, one for tasks 7 to 11 and one for tasks 12 to 16. Division of the task order into three groups allows for the possibility of an increasing, decreasing or U-shaped pattern of scale with respect to order.

In the event, our concerns regarding the confounding impact of the experimental design are borne out by the data. Observe from Table 1 that the scale parameters on tasks 1 and 17, are similar in magnitude but considerably smaller than the three parameters tracing the development of the scale parameter over the middle 15 choice tasks. A likelihood ratio test reveals that we can categorically reject the hypothesis that the scale parameter is constant across all 17 choice

³² An important practical reason for controlling for possible position-dependent changes in the scale parameter is that preference parameters are only estimated up to scale so that failing to properly control for changing error scale can lead to the erroneous identification of position-dependent order effects in the preference parameters.

tasks (χ^2 statistic = 29.274, df = 4, p -value: <0.001). In all probability that significant difference reflects the comparatively small variance in attribute levels inherent in the first and last tasks. In contrast, we cannot reject the hypothesis that the scale parameter takes the same value for the identical first and last tasks as well as taking the same (but different) value across the middle 15 choice tasks (χ^2 statistic = 3.878, df = 3, p -value: 0.275). Having controlled for design-induced differences in the scale parameter, our data do not provide evidence of either learning (preference or institutional) effects, in the sense of responses becoming less variable as respondents progress through the sequence of choice tasks, or fatigue or failing concentration effects, in the sense of increasingly variable responses in later tasks.³³

Precedent-Dependent Ordering Effects

The inclusion of three precedent variables in the model allow us to test for the possibility that individuals responses are somehow framed by the nature of other options observed in the SP exercise. The first of those precedent variables is the worse deal indicator S_j which takes a value of 1 when the option under consideration offers a deal that is worse than some previously observed deal and 0 otherwise (see equation 11). The parameter estimated on the worse deal indicator, γ_1 , is negative and highly significant. Respondents, it appears, are significantly more likely to refuse a particular option when they have previously been offered a better option. That pattern of behavior is strongly suggestive of strategic misrepresentation of preferences.

The second precedent variable is the starting point comparison variable Q_j . That variable records the difference between the deal offered by the current task and that offered by the initial task (see equation 12). Positive values indicate a better deal in the current task, negative values the reverse. The parameter estimated on the starting point comparison variable, γ_2 , is positive and highly significant. Our data suggest that individuals respond more favorably to a particular option if it offers a deal that is superior to the deal they happened to have been offered in the initial task. Conversely they respond less favorably to that same option when it is inferior to the deal they happen to have been presented with in the first task. That behavior amounts to what is generally termed a starting point effect the diverse interpretations of which were discussed previously.

³³Of course, if learning or fatigue effects of this ilk occur almost exclusively between the first and the second choice tasks, then such changes would be masked in our data by the confounding impact of the changing attribute variance exhibited by options in those two choice tasks. Given the similarity in the scale parameter for the first and last task, however, that possibility seems remote. As a result, any such effect working between 1st and 2nd choice tasks would have to have been offset by a similarly sized but oppositely signed change between the 16th and 17th tasks.

Our final precedent variable is that constructed to test for the possibility of behavior resembling price referencing. The reference deal comparison variable, R_j (see equation 14), captures the difference between the deal in the current option and the reference deal, where the reference deal is some suitably constructed average of the deals offered in preceding tasks. The parameter η dictates whether all previous deals should be weighted equally in the calculation of that average or whether more weight should be attached to more recently or, alternatively, more distantly observed deals. Observe from Table 1 that the parameter estimated on the reference deal comparison variable, γ_2 , is positive and highly significant. It appears that individuals are more likely to respond positively to a particular option if the deal it offers is perceived as a gain relative to the reference deal than when it is perceived as a loss. Notice also that the weighting parameter η is estimated to have a value of -1.429 , indicating that the reference deal to which individuals draw comparison is shaped primarily by the deals observed in recently observed tasks. Given the functional form for calculating the reference deal (equation 13), the weights attached to deals in previous tasks vary depending on order in the sequence; clearly, in the second task all weight must be attached to the deal seen in the first task, but by the final task some weight is attached to all 16 preceding tasks. On average across the 17 tasks, however, our model suggests that the deal seen in the very last task counts for 55% in the calculation of the precedent. Likewise the last 3 tasks account for 82% and the last 5 tasks for 91%. Accordingly, our data provide strong support for the reference price hypothesis. As far as we are aware, this is the first time that the literature has documented evidence for individuals developing reference deals from their exposure to choice tasks and using those reference deals in order to frame their responses.³⁴

Let us return briefly to the consideration of the highly significant order effect documented in our non-parametric analysis of the repeated first and last tasks. There we observed a distinct reduction in respondents' propensity to choose the alternative option in the last task when compared to the first (see Figure 3). Our econometric analysis provides some insights as to how preferences evolve across the series of choice tasks in order to occasion that observed empirical regularity. First, as shown in Figure 4, preferences are subject to a position-dependent development of the status quo effect. While the first task shows very little evidence of a status quo effect by the seventeenth task respondents are significantly more likely to reject the alternative option in favor of the status quo. In addition, our econometric model reveals a number of precedent-dependent order effects. Of those, we have no reasons to suspect that

³⁴ Though one could reinterpret the 'context dependence' effect observed by Holmes and Boyle (2005) as reference pricing and Hu (2007) finds evidence that respondents use experience of prices from real markets as reference prices in answering SP questions for a market good (canola oil).

starting point or reference-price effects should result in any systematic difference in preferences between the first and last task. In contrast, strategic misrepresentation of preferences provides a supporting explanation for the first task - last task disparity. Our econometric model suggests that respondents are more likely to reject options that present them with deals worse than the best previous deal they have seen. In the first task respondents have, of course, not been exposed to previous deals such that responses in that task cannot be subject to this form of strategic misrepresentation. Only in the event that the first task (and hence the last task also) offers the best deal in the whole sequence of tasks will that also be true of responses to the last task. In all other cases, respondents will view an intervening task offering a deal better than the first and final task. Accordingly, when it comes to responding to that last task, a strategic respondent has incentives to misrepresent their preferences by opting for the status quo. Again, last task responses are likely to show lower demand for the alternative option than first task responses.

Finally, it is worthy of note that our results indicate that the response to the first question has some interesting and potentially appealing properties. Specifically the status quo effect is virtually zero, strategic behavior must be at most minimal and we can say nothing about starting point effects. Some might see these conditions as sufficient to argue for placing special emphasis upon this first response; an argument which accords with that of Carson and Groves (2007).

7. Concluding observations

The intended contribution of this paper was to provide a framework for detection and separate identification of differing forms of ordering effect within repeated response stated preference studies; effects which challenge the task independence assumption underlying such studies.

We start from first principles with the standard structural model of choice. This is extended to allow detection of signature patterns of response which separately identify two general forms of failure in the task independence assumption: position dependence effects (specifically preference learning, institutional learning and fatigue); and precedent dependence effects (specifically starting point effects, full recall strategic behavior and reference price behavior).

In designing an empirical test of this extended structural model for detecting ordering effects our application seeks to avoid various potential confounding effects. We note that previous studies have typically tested for order related issues such as fatigue or learning by examining patterns within error variance as respondents progress through a question sequence.

However, such tests may be confounded where the mix of choice sets across orders does not share similar covariance matrix elements. We address this through a Latin square design intended to generate a more balanced mix of choice sets across the question order. A further potential problem in the form of misspecification bias was addressed by employing a full factorial design allowing estimation of both main and interaction effects independently of one another. In addition that empirical analysis adopted a data-driven approach to the specification of the functional form of the various ordering effects under analysis.

It seems reasonable to suppose that certain ordering effects might be case study specific. For example, issues such as fatigue might depend upon the complexity of the good under investigation and the cognitive demands imposed by the chosen choice tasks presented to respondents. For this primarily methodological study we deliberately chose a relatively simple good, readily described by price and two other attributes. Empirical findings might therefore be specific to this case and are thus only briefly reviewed. We find no evidence of fatigue and preference learning. However, signature patterns suggest that respondents do shift stated preferences early on in the question sequence and that all three precedent-dependent ordering effects (starting point, strategic behaviors and price referencing) are evident in these data.

We believe that the design principles of this study are of more general interest. The needs for control against induced changes in error variance arising from the design and allocation of choice tasks to a question ordering is, we believe, of general importance. Similarly the clear significance of interaction effects reported here may provide some caveat to the widespread use of main effect designs. Most generally we believe that the use of a structural model of choice to underpin our tests provides an approach to future testing which we suggest may be of lasting usefulness.

Finally, combining these general messages and specific findings we suggest that a fruitful line for future research might be to apply the approach proposed here across a wider array of goods and choice task designs. This would permit examination of the generality of our specific findings across key real world dimensions and begin to build up a picture of the incidence and importance of issues such as strategic behavior and other forms of order related phenomena thereby improving the information available to decision makers.

References

- 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.
- W.L. Adamowicz, J.J. Louviere, M. Williams, Combining revealed and stated preference for valuing environmental amenities, *Journal of Environmental Economics and Management* 26 (1994) 271-292.
- A. Alberini, B. Kanninen, R.T. Carson, Modeling response incentives in dichotomous choice contingent valuation studies. *Land Economics* 73(1997) 309-324.
- J. Alevy, C. Landry, and J. List, Anchors away: field experiments on anchoring of economic valuations, East Carolina University 2008 Working Papers, ecu0810, July 2008.
- O. Amir, J. Levav, Choice construction versus preference construction: the instability of preferences learned in context, *Journal of Marketing Research* 45 (2008) 145–158.
- K. Arrow, R. Solow, P.R. Portney, E.E. Leamer, R. Radner and H. Schuman, Report of the NOAA Panel on Contingent Valuation, *Federal Register* 58 (1994) 4601-4614.
- 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 (2008a) 127-141.
- I.J. Bateman, Carson, R.T., Day, B.H., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and -Wang, P., Choice Set Awareness and Ordering Effects in Choice Experiments, CSERGE Working Paper EDM 08-01, Centre for Social and Economic Research on the Global Environment, University of East Anglia, (2008b).
- I.J. Bateman, R.T. Carson, B.H. Day, W.M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Özdemiroğlu, D.W. Pearce, R. Sugden, J. Swanson, *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar Publishing, Cheltenham, 2002.
- I.J. Bateman, I.H. Langford, A.P. Jones, and G.N. Kerr, Bound and path effects in double and triple bounded dichotomous choice contingent valuation. *Resource and Energy Economics* 23 (2001) 191-213.
- K.J. Boyle, R.C. Bishop, M.P. Welsh, Starting point bias in contingent valuation bidding games, *Land Economics* 61 (1985) 188–194.
- J. Braga, C. Starmer, Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis, *Environmental and Resource Economics* 32 (2005) 55-89.
- J. Brazell, J. Louviere, Respondent's help, learning and fatigue. Presented at the 1997 INFORMS Marketing Science Conference, University of California, Berkeley, 1997.
- R.A. Briesch, R.A., L. Krishnamurthi, T. Mazamdur, S.P. Raj, A comparative analysis of reference pricing models, *Journal of Consumer Research* 24 (1997) 202-214.
- T. Brown, D. Kingsley, G. Peterson, N.E. Flores, A. Clarke, and A. Birjulin, Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement, *Journal of Public Economics* 92(2008) 1595–1606.
- R.E. Bucklin, R.E. Lattin, Reference effects of price and promotion on brand choice, *Journal of Marketing Research* 26 (1989) 299-310.
- T.A. Cameron, T.A. and 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.
- F. Carlsson, F., P. Martinsson, How much is too much?—An investigation of the effect of the number of choice sets, starting point and the choice bid vectors in choice experiments, *Environmental and Resource Economics* 40 (2008) 165–176.
- R.T. Carson, *Contingent Valuation: A Comprehensive Bibliography and History*, Edward Elgar, Northampton, MA, forthcoming.
- R.T. Carson, T. Groves, Incentive and Informational Properties of Preference Questions, *Environmental and Resource Economics* 37 (2007)181-210.
- J.R. Corrigan, C.L. Kling, J. Zhao, Dynamic willingness to pay: An empirical specification and test, *Environmental and Resource Economics* 40 (2008) 285-298.
- S. Caussade, S., J. de D. Ortuzar, L. Rizzi, D. Hensher, Assessing the Influence of Design Dimensions on Stated Choice Experiment Estimates. *Transportation Research B*, 39 (2005) 621–640.
- P.A. Champ, K. Boyle, T.C. Brown (eds.), *A Primer on Non-market Valuation*, Kluwer Academic Press, Dordrecht, 2003.

- Y.L. Chien, C.J. Huang, D. Shaw, A general model of starting point bias in double-bounded dichotomous contingent valuation surveys, *Journal of Environmental Economics and Management* 50 (2005) 362–377.
- B.H. Day, J.L. Pinto, Ordering anomalies in choice experiments, Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, 2009.
- Defra (Department for Environment, Food and Rural Affairs), An introductory guide to valuing ecosystem services, Defra, London, available at www.defra.gov.uk, 2007.
- B.G. Dellaert, J.D. Brazell, J.J. Louviere (1999), The effect of attribute variation on consumer choice consistency, *Marketing Letters* 10 (1999) 139-147.
- 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.
- J.R. DeShazo, Designing transactions without framing effects in iterative question formats. *Journal of Environmental Economics and Management* 43 (2002) 360-385.
- 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.
- A. Gabor, C.W.J. Granger, On the price consciousness of consumers, *Applied Statistics* 10 (1961) 170-188.
- W.M. Hanemann, J. Loomis, B. Kanninen, Statistical efficiency of double-bounded dichotomous choice contingent valuation, *American Journal of Agricultural Economics* 73 (1991) 1255-1263.
- D. Hensher and J. Rose, Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay, *Transportation Research Part E* 45 (2009) 583–590.
- 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.
- I.P. Holman, P.J. Loveland, R.J. Nicholls, S. Shackley, P.M. Berry, M.D.A. Rounsevell, E. Audsley, P.A. Harrison, R. Wood, REGIS: regional climate change impact response studies in East Anglia and North West England,” report to UK Defra, 2002.
- 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.
- T.P. Holmes, R.A. Kramer, An independent sample test for yea-saying and starting point bias in dichotomous choice contingent valuation, *Journal of Environmental Economics and Management* 29 (1995) 121–132.
- W. Hu, Bargains or rip-offs: reference price effects in conjoint demand, *Journal of Agricultural and Resource Economics* 32 (2007) 256-272.
- P.R. Hunter, Climate change and waterborne and vector-borne disease, *Journal of Applied Microbiology* 94 (2003) 37-46.
- 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.
- P. Kabat, R.E. Schulze, M.E. Hellmuth, J.A. Veraart, eds., *Coping with impacts of climate variability and climate change in water management: a scoping paper*, DWC-Report no. DWCSSO-01(2002), International Secretariat of the Dialogue on Water and Climate, Wageningen, Netherlands, 2002.
- G. Kalyanaram, R.S. Winer, Empirical generalizations from reference price research, *Marketing Science*, 14 (1995) G161-G169.
- 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.
- J.J. Louviere, D. Hensher, J. Swait, *Stated choice methods: analysis and application*, Cambridge: Cambridge University Press, 2000.
- R. D. Luce, Thurstone and sensory scaling: then and now, *Psychological Review* 101 (1994) 271-277.
- T. Maddala, K.A. Phillips, F.R. Johnson, An experiment on simplifying conjoint analysis designs for measuring preferences. *Health Economics*, 12 (2003) 1035-1047.
- G.E. Mayhew, R.S. Winer, An Empirical Analysis of Internal and External Reference Prices Using Scanner Data, *Journal of Consumer Research*, 19 (1992) 62-70.
- T. Mazumdar, S.P. Raj, and I. Sinha, Reference price research: review and propositions, *Journal of Marketing* 69 (2005) 84-102.

- D. McFadden, Conditional logit analysis of qualitative choice analysis, in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 1974.
- R. Mitchell, R. Carson, *Using surveys to value public goods: the contingent valuation method*, Resources for the Future, Washington, DC, 1989.
- MORI, *The 2004 Periodic Review: research into customers' views*, London: DEFRA, 2002.
- R. O'Connor, M. Johannesson, P.O. Johansson, Stated preferences, real behavior and anchoring: some empirical evidence, *Environmental and Resource Economics* 13 (1999) 235–248.
- T. Ohler, A. Le., J. Louviere, J. Swait, Attribute range effects in binary response tasks, *Marketing Letters* 11 (2000) 249–260.
- C.R. Plott, Rational individual behavior in markets and social choice processes: the discovered preference hypothesis, in *Rational foundations of economic behavior*, K. Arrow, E. Colombatto, M. Perleman, C. Schmidt, eds., London: Macmillan, 1996
- Putler, D.S., Incorporating reference price effects into a theory of consumer choice, *Marketing Science* 11 (1992) 287–309.
- M. Rothschild, Models of market organization with imperfect information, *Journal of Political Economy* 81(1973) 1283–1308.
- R. Rowe, R. C. d'Arge, D. S. Brookshire, An experiment on the economic value of visibility, *Journal of Environmental Economics and Management* 7 (1980) 1–19.
- S. Salop, J. Stiglitz, Bargains and rip-offs: a theory of price dispersion with monopolistic competition, *Review of Economic Studies* 44 (1977) 493–510.
- S. Savage, D. Waldman Learning and fatigue during choice experiments: a comparison of online and mail survey modes, *Journal of Applied Econometrics* 23 (2008)351–371.
- R. Scarpa, E.S.K. Ruto, P. Kristjanson, M. Radeny, A. Drucker, J.E.O. Rege, Valuing indigenous cattle breeds in Kenya: an empirical comparison of stated and revealed Preference value estimates, *Ecological Economics* 45 (2003) 409–426.
- R. Slonim, E. Garbarino Similarities and differences in stockpiling and reference effects, *Managerial and Decision Economics* 30 (2009) 351–371.
- D. Stahl, Oligopolistic pricing with sequential consumer search. *American Economic Review* 79 (1989) 700–712
- A.P. Street, D.J. Street, *Combinatorics of experimental design* Oxford: Oxford University Press, 1986.
- G. Stigler, The economics of information, *Journal of Political Economy* 69 (1961) 213–229.
- 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.
- L.O. Taylor, M.D. Morrison, K.J. Boyle, Exchange rules and the incentive compatibility of choice experiments, manuscript, Department of Agricultural and Resource Economics, North Carolina State University, 2007.
- R. Thaler, Mental accounting and consumer choice, *Marketing Science* 4 (1985) 199–214.
- L. Thurstone, A law of comparative judgment, *Psychological Review* 34 (1927) 273–286.
- A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, *Science* 185 (1974) 1124–1131.
- A. Tversky, D. Kahneman, Loss aversion in riskless choice: a reference-dependent model, *Quarterly Journal of Economics* 106 (1991) 1039–1061.
- J. Whitehead, Incentive incompatibility and starting-point bias in iterative valuation questions, *Land Economics* 78 (2002) 285–297.
- P. Wicksteed, *The common sense of political economy*, Volume I. Routledge, London, 1910.

Figure 1: Typical choice question

	'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

Choose

'No Scheme'

'Scheme A'

Which would you choose?

(tick one box only)

Figure 2: Sample proportions choosing the alternative option at different costs in the first and last tasks for the larger as compared to smaller water quality improvement

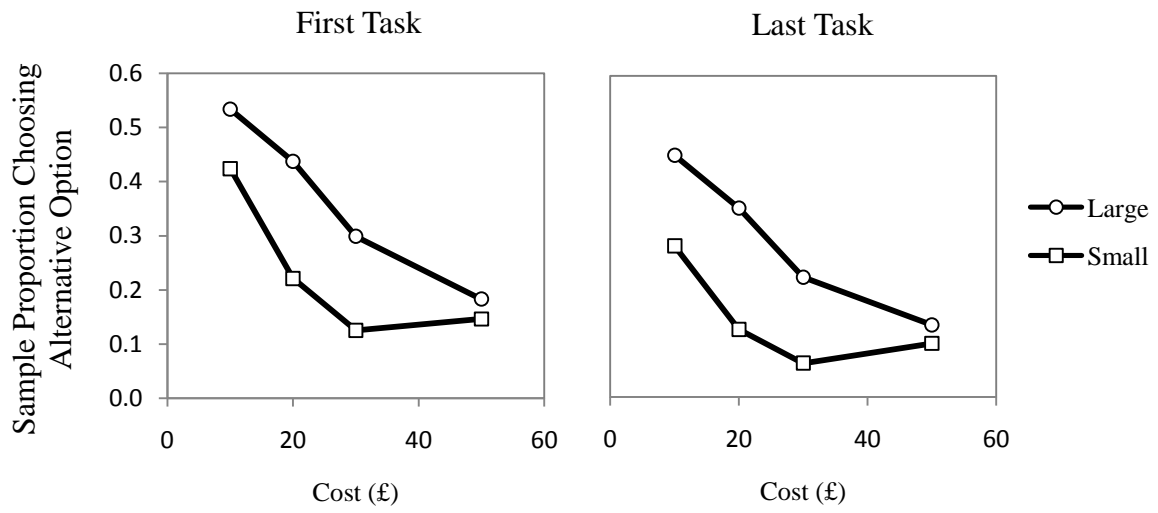


Figure 3: Sample proportions choosing the alternative option at different costs for the larger and smaller water quality improvements in the first as compared to last task

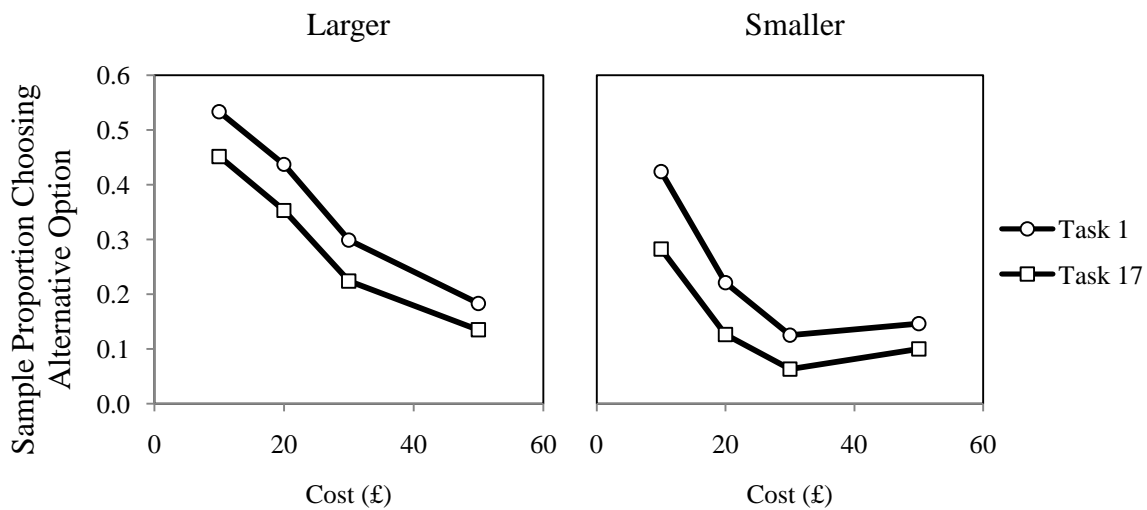


Figure 4: Position-dependent ordering effect in the status quo effect parameter

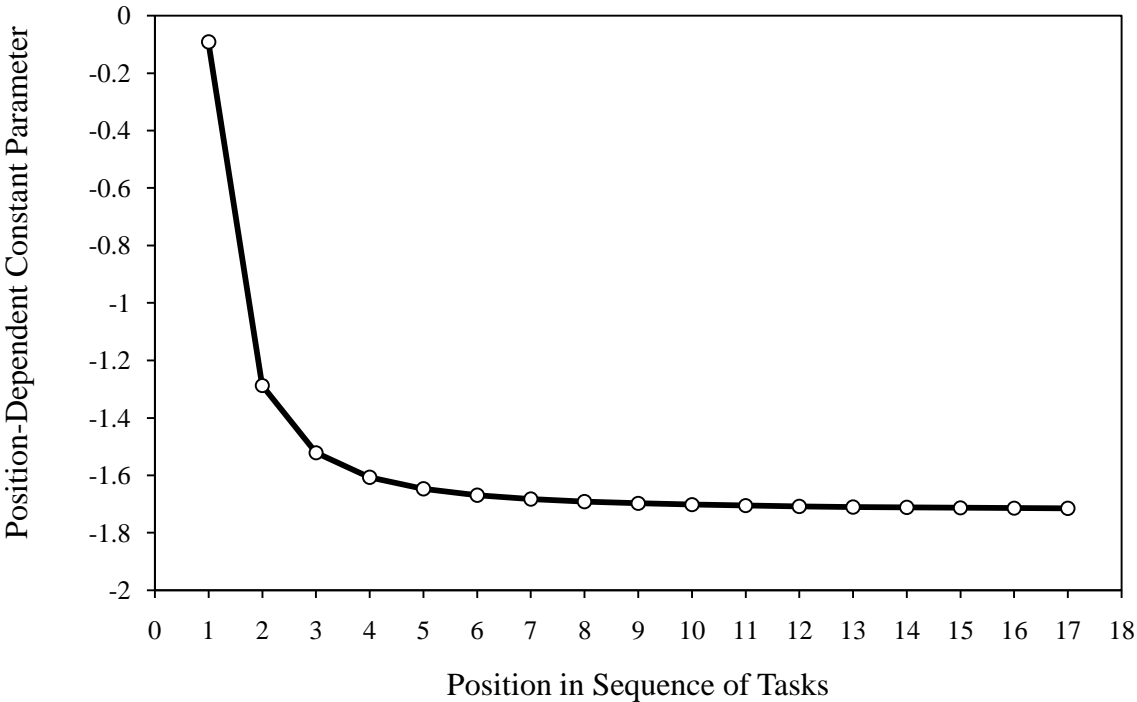


Table 1: Model parameters from random effects probit estimation

Variable	Parameter Estimate	Std. Err.	<i>t</i> -stat	<i>p</i> -value
<i>Base Preference Parameters:</i>				
Constant (α)	-1.7225	0.1341	-12.845	<0.001
<i>Main Effects:</i>				
Odor (β_1)	-0.1205	0.0094	-12.822	<0.001
Color (β_2)	-0.2710	0.0182	-14.893	<0.001
Cost (β_3)	-0.0429	0.0021	-20.078	<0.001
<i>Interaction Effects:</i>				
Color \times Odor (γ_1)	-0.0069	0.0027	-2.541	0.011
Cost \times Color (γ_2)	0.0021	0.0005	4.133	<0.001
Cost \times Odor (γ_3)	0.0008	0.0003	2.717	0.007
<i>Position-Dependent Order Effects:</i>				
<i>In Preference Parameters:</i>				
Trend \times Odor (b_1)	0.1856	0.1734	1.070	0.285
Trend \times Color (b_2)	-0.1160	0.2701	-0.430	0.668
Trend \times Cost (b_3)	-0.0062	0.0044	-1.396	0.163
Power Weight in Trend (λ_a)	-10.8375	28.3476	-0.382	0.702
<i>In Constant:</i>				
Trend \times Constant (a)	1.6304	0.6736	2.420	0.016
Power Weight in Trend (λ_b)	-1.9071	0.5678	-3.359	0.001
<i>Precedent-Dependent Order Effects:</i>				
Worse Deal Indicator (θ_1)	-0.1538	0.0525	-2.929	0.003
Starting Point Comparison (θ_2)	0.0455	0.0079	5.751	<0.001
Reference Deal Comparison (θ_3)	0.0158	0.0057	2.794	0.005
Power Weight in Ref Deal (η)	-1.4288	0.5988	-2.386	0.017
<i>Scale Parameters:</i>				
σ_1	0.5306	0.0556	9.545	<0.001
σ_2 to σ_6	0.7034	0.0325	21.635	<0.001
σ_7 to σ_{11}	0.7428	0.0354	20.987	<0.001
σ_{12} to σ_{16}	0.6863	0.0315	21.780	<0.001
σ_{17}	0.4312	0.0458	9.408	<0.001
<i>Respondents</i>	833			
<i>Observations</i>	14,134			
<i>lnL</i>	-4,357.013			