CBESS Discussion Paper 17-06

Attention and Novelty: An Experimental Investigation of Order Effects in Multiple Valuation Tasks

by Cameron A. Belton *

* School of Economics and CBESS, University of East Anglia

Abstract
This paper implements a novel experimental design to investigate the presence of order effects across multiple valuation tasks for consumer goods, whereby earlier goods are valued more highly than later goods. The paper presents a novel explanation of order effects, relating to attention and novelty. Typically, multiple valuation tasks for consumer goods use the same good for valuation in each task. In this experiment the type of good valued in each task is varied, allowing two potential mechanisms to be disentangled: experimental novelty effects, whereby participants become less interested with completing later tasks, and good-specific novelty effects, whereby participants become less interested with the goods used in later tasks. The results find a particular importance of the first task; goods in the first task are valued significantly higher than later valued goods, evidence of experimental novelty effects, and goods of a similar type to the good in the first task are valued significantly higher than goods of a different type to the first, evidence of good-specific novelty. The paper discusses the potential implications of these findings.

JEL classification codes
C91, D12, D91

Keywords
order effects; attention; novelty; advanced disclosure

University of East Anglia

Centre for Behavioural and Experimental Social Science

Centre for Behavioural and
Experimental Social Science
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/cbess

Attention and Novelty: An Experimental Investigation of Order Effects in Multiple Valuation Tasks*

Cameron A. Belton [a]

[a] Centre for Behavioural and Experimental Social Science, School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK

July 31th, 2017

Abstract

This paper implements a novel experimental design to investigate the presence of order effects across multiple valuation tasks for consumer goods, whereby earlier goods are valued more highly than later goods. The paper presents a novel explanation of order effects, relating to attention and novelty. Typically, multiple valuation tasks for consumer goods use the same good for valuation in each task. In this experiment the type of good valued in each task is varied, allowing two potential mechanisms to be disentangled: experimental novelty effects, whereby participants become less interested with completing later tasks, and good-specific novelty effects, whereby participants become less interested with the goods used in later tasks. The results find a particular importance of the first task; goods in the first task are valued significantly higher than later valued goods, evidence of experimental novelty effects, and goods of a similar type to the good in the first task are valued significantly higher than goods of a different type to the first, evidence of good-specific novelty. The paper discusses the potential implications of these findings.

JEL Classification: C91, D12, D91

Keywords: order effects; attention; novelty; advanced disclosure

*Corresponding author: Cameron Belton (cameron.belton@uea.ac.uk)

1. Introduction

In experimental economics and in stated preference studies, it is common to use designs that generate multiple responses from individuals. Multiple responses are an intrinsic part of within-subject experimental designs where the aim is to compare individuals' responses to different tasks. They are also essential for the methodology of discrete choice experiments, which elicit many binary decisions from each respondent as a means of estimating individuals' valuations of non-marketed goods or services. Multiple-response designs are also a useful way of reducing the costs of experimentation by generating multiple data points from each participant. However, there is a growing literature which suggests that multiple-response designs are vulnerable to *order effects*: the order in which tasks are presented can affect the outcomes of these decisions. These effects are particularly problematic when, as in stated preference studies, the aim is to estimate the distribution of preferences in some population. Order effects have been found in a number of stated preference methodologies, including choice experiments for non-marketed goods (Day and Pinto Prades, 2010; Day *et al*., 2012), contingent valuation surveys for hypothetical environmental goods (Bateman *et al*., 2004; Payne *et al*., 2000), and multiple-task valuation exercises for real goods (Ariely *et al*., 2003; Clark and Friesen, 2008). Order effects pose a less direct problem when within-subject designs are used to make qualitative comparisons between behaviour in different experimental treatments, since in these cases, counterbalancing of the order of tasks can be used as a control. Even so, the presence of order effects casts doubt on whether individuals possess stable preferences that can be elicited through surveys or experiments. Understanding why order effects occur is a fundamental problem for experimental economics.

The existing literature has identified a wide range of mechanisms by which, when an individual faces a sequence of valuation elicitation tasks, earlier tasks may induce the formation of reference points which then affect responses to later tasks. Such mechanisms include anchoring effects (Ariely *et al*. 2003), embedding (or part-whole) effects (Kahneman and Knetsch, 1992; Bateman *et al*. 1997), and shaping effects (Loomes *et al*. 2003). I will refer to these effects as *indirect* order effects.

This paper looks at order effects from a new perspective. It investigates the possibility that there is a more direct causal mechanism which induces *attention-based* order effects. Such an effect would be one which causes a systematic trend in responses as task order progresses, simply as a result of the reduced attention that participants give to later tasks. Specifically, this paper tests for evidence of reduced attention generating a systematic decrease in valuations across multiple selling valuation tasks, for different real, consumable goods.

Previous experimental studies have found systematic declines in the valuations of sellers across repeated valuation tasks for the same good (Shogren *et al*., 1994; Loomes *et al*., 2010). Loomes *et al*. (2010, p. 385) suggest this may be a result of a reduced sense of loss aversion. Their hypothesis is that sellers are reluctant to give up a good in earlier tasks (generating relatively high selling valuations), but that repeating the selling task for the same good dulls the sense of loss aversion in later tasks and so reduces valuation, as the salience of not selling weakens.

This paper argues that such a decline may be a consequence of the reduced attention that participants give to later tasks. Previous economic research has found that an increase in attention toward the positive attributes of a good can result in an increase in valuations (Carmon and Ariely, 2000; Nayakankuppam and Mishra, 2005). From a psychological perspective, attention toward a novel stimulus diminishes over time (Berlyne, 1951). While being asked to value goods or services in earlier tasks is originally a novel exercise for participants, as task order progresses, such novelty value falls.

These findings suggest that, as novelty diminishes, so too does the attention given to that less novel stimulus, either as a result of familiarity with the act of completing the task itself, or as the specific details within the task become less novel, or a combination of the two. The role of diminishing attention is therefore an important one for experimental methodology, as it implies that participants may think less or more about tasks depending on the order in which they complete them.

In previous experiments that have found systematic declines in the valuations of sellers across repeated valuation tasks, the market or non-market goods (Shogren *et al*., 1994) or lotteries (Loomes *et al*., 2010) were identical in every task. In such a design it is not possible to determine whether the decline in valuations is a result of reduced attention to the act of completing tasks in general or to the particular good that is being valued, since both are repeated identically in each task. The experiment reported in the present paper attempts to disentangle these effects by manipulating the type of goods valued in each task in such a way that the effects of different kinds of novelty can be separated.

In the experiment, each of the six selling tasks faced by each participant involved a different good. The goods used in the experiment were chosen to allow the effects of different degrees of novelty to be investigated. The six goods can be partitioned into two subsets, one containing three distinct but similar patterned mugs, the other containing three distinct but similar luxury chocolate bars. In this design, changes in attention induced by the relative novelty of tasks and goods may generate order effects in three ways, each of which can be investigated in isolation. A task-specific *experimental novelty effect* may occur as a result of reduced attention to the act of completing later tasks, independent of the specific goods used. Two kinds of good-specific novelty effects are possible. A *within-subset novelty effect* may occur if participants give reduced attention to goods that are similar to goods that have featured in previous tasks. A *between-subset novelty effect* may occur if participants give more attention to the first type of good (mugs or chocolates) they confront than to the second.

The results of the experiment show a general tendency for a decline in valuations over the six selling tasks. This decline is especially strong immediately after the first task – evidence of an experimental novelty effect. There is also evidence that valuations are higher for the first type of good that a participant confronts than for the second – consistent with a between-subset novelty effect. These findings suggest that both the general completion of tasks and the specific goods used for valuation may be responsible in generating attention-based order effects.

The rest of the paper proceeds as follows. Section 2 discusses current theories of indirect order effects, and the role of attention as a possible cause of order effects. Section 3 describes the experimental design and Section 4 outlines the key hypotheses to be investigated. Section 5 presents the results and Section 6 provides a discussion of the implications of these findings. Section 7 concludes.

2. Explanations of Order Effects

2.1 Indirect Order Effects

The persistence of order effects across different stated preference methodologies that have been used to elicit valuations of marketed and non-marketed goods has resulted in a number of theories explaining this occurrence. (For a detailed review of these theories, see Day *et al*., 2012.)

Anchoring effects, whereby responses to earlier questions are used as cues to shape responses to later ones, have been observed in stated preference exercises as starting point bias (for example, Herriges and Shogren, 1996). The impact of anchoring has been shown with even arbitrary anchors, such as a spin of a roulette wheel (Tversky and Kahneman, 1974) or respondents' social security numbers (Ariely *et al*., 2003), influencing consequent and unrelated decision-making exercises. When comparing the differing effects of arbitrary anchoring on buying and selling prices, existing research presents conflicting accounts, finding that selling prices may be more (Sugden *et al*., 2013) or less (Simonson and Drolet, 2004) affected by anchoring than buyers, under differing conditions.

When respondents are asked to value bundles of goods and when the bundles involved in different questions differ in size or objective value, embedding effects (or part-whole bias) may occur. For example, if one question elicits a valuation of the set of goods {A, B, C} while another elicits a valuation of {A, B}, the valuation of {A, B} tends to be higher if it elicited before that of {A, B, C}. One possible explanation of embedding in hypothetical contingent valuation surveys is that valuation is a proxy of purchase of moral satisfaction (Kahneman and Knetsch, 1992), suggesting that the moral satisfaction of contributing doesn't depend on what is actually achieved. However, research on nested bundles of objectively differing private goods provides evidence of embedding effects, even with incentivised valuation elicitation devices (Bateman *et al*., 1997; Ariely *et al*., 2003; Clark and Friesen, 2008), suggesting that the purchase of moral satisfaction is not a complete explanation of embedding effects.

Shaping effects may induce order effects in certain kinds of experimental market. When valuations of a given good are elicited in repeated markets, the valuations that are implicit in participants' bids to buy or sell tend to move towards prices set in previous markets (Loomes *et al*, 2003), possibly because participants are unsure about the true value of the good to them. In a repeated Vickrey second price auction (such as in Shogren *et al*., 2001), the selling price is equal to the second lowest bid, and so shaping effects can induce downward trends in sellers' valuations. Knetsch *et al*. (2001) find evidence that manipulating the design of a Vickrey second price auction may generate different shaping effects (potentially eliminating any decay in valuations), and so reveal concerns about the demand-revealing properties of these auction designs.

In some instances, order effects may be generated by strategic bidding which diminishes over time (Shogren *et al*., 1994, p.266). In surveys which elicit valuations for non-marketed commodities, repeated questions involving different levels or costs of public good provision may reduce the credibility of any given level or cost being actually implemented (Carson and Groves, 2007, p.185). Fatigue effects have also been shown to present potential issues for stated preference choice experiments (e.g. Savage and Waldman, 2008), creating tendencies for favouring the status quo or increased randomness in responses (Day *et al*., 2012, p.75).

There is evidence from psychological experiments that when participants carry out a sequence of tasks, each of which requires them to rate the desirability of pairs of trivially differing images, the first viewed images tend to be judged more favourably than later ones (Pandelaere *et al*., 2010). Possible explanations for this finding include the strength of memory retrieval of first goods (Bruce and Papay, 1970), the perceived legitimacy of first goods as the 'original', or the positive mental imprint first viewed goods leave relative to consequent goods (Pandelaere *et al*., 2010, pp.447-448).

2.2 Attention-Based Order Effects

In studies of the effects of experience on the well-known disparity between willingness-to-pay of buyers and willingness-to-accept of sellers, it is common to find a systematic decline in the valuations of sellers, but not in those of buyers (e.g. Shogren *et al*., 1994; Loomes *et al*., 2010). As noted in Section 1, the decline in sellers' valuations might be the result of reductions in loss aversion. Loss aversion with respect to the good would be irrelevant for buyers.[1]

Attention-based order effects offer an alternative explanation of why such a decline in valuations is found for sellers but not buyers. Theories of attentional bias (Carmon and Ariely, 2000) suggest that individuals tend to give more attention to what they stand to forgo in a transaction than to what they stand to gain. Since sellers stand to forgo a good, it is sellers who focus more on the attributes of goods. If the more an individual focuses their attention on the (positive) attributes of a good, the more desirable the good is perceived to be, then this may lead to higher valuations by sellers than by buyers. If attention falls as task order progresses, then it follows that later goods are perceived as less desirable, and so their valuations decline.

Research from psychology suggests that attention is also positively related to the novelty value of a stimulus (Berlyne, 1951; Betsch *et al*. 1998). The potential significance of novelty for economics was noticed by Scitovsky (1992) but, to date, there has been relatively little investigation of the role of novelty in economic decision making. There is some evidence that novelty increases the appeal of certain kinds of goods (Tom, 2004; Tom *et al*., 2007) and is a cause of increased willingness-to-pay for new variants of recognisable foods (Jaeger and

[1] If buyers' valuations were affected by loss aversion with respect to money, declining loss aversion would induce an increase in those valuations. Whether loss aversion for money has quantitatively significant effects on decisions involving low-value goods is an unresolved question (Bateman et al., 2005).

Harker, 2005; Stevens and Winter-Nelson, 2008; Meenakshi *et al*. 2012), suggesting it can be responsible for tangible effects on the way that goods are perceived.

There are at least three mechanisms by which variations in novelty, and corresponding variations in attention, might generate order effects in multiple-task surveys and experiments which elicit selling valuations.

The first mechanism is an experimental novelty effect – the reduced novelty of the specific experimental methodology as an experiment or survey progresses. With typical lab experiments using student participants, or contingent valuation experiments using telephone or face-to-face interviews with members of the public, the average participant is unlikely to have taken part in many multiple-task studies previously. Thus, earlier tasks are undoubtedly more novel to participants, but this novelty may quickly dissipate once participants assimilate to the nature of the tasks. The greater novelty of earlier tasks might induce more attention and thereby (because of attentional bias towards what is forgone) higher selling valuations. This mechanism would tend to induce a general decline in selling valuations, independent of changes in the specifics of the task.

In an experiment eliciting multiple valuations of hypothetical environmental goods, a *primacy effect* has been found: the reduction in valuation between the first and second valuation task is greater than that between other adjacent tasks (Payne *et al*., 2000). That experiment elicited willingness-to-pay for public contributions to a variety of environmental goods, but if experimental novelty effects do drive declines in valuations across tasks, it may be that primacy effects are even stronger in selling tasks.

The second mechanism is a within-subset novelty effect. If the goods valued in different tasks are similar to one another, then the novelty of the good that a participants is valuing will fall as the number of similar goods previously valued increases. This mechanism would tend to induce a decline in selling valuations within any given subset of similar goods.

The third mechanism is a between-subset novelty effect. If the goods in an experiment are naturally thought of as belonging to distinct types that differ substantially from one another, then the introduction of a new type of good may be perceived as a shift from one type of task (for example, ‘valuing chocolates’) to another type (‘valuing mugs’). Since the second type of good to be introduced would no longer be associated with the initial novelty of the experiment, this mechanism would tend to induce lower selling valuations for later types of good than for earlier types.

In contingent valuation studies, one attempt to rectify issues of order is to employ a method of advanced disclosure whereby survey participants are fully informed as to the type of goods to be valued before any valuations are elicited. This procedure has been found to dissipate some order effects (Bateman *et al*., 2004; Day *et al*., 2012). If good-specific (within- or between-subset) novelty effects were driving a downward trend in valuations, then it is possible that advanced disclosure would reduce this effect, as the initial information about the goods to be valued might remove any good-specific novelty that would otherwise be revealed at the start of each new valuation task.

3. Experimental Design

The experiment reported in this paper was designed to test for attention-based order effects when valuations are elicited in selling tasks. An important feature of the design is its ability to isolate the different attention-based mechanisms described in Section 2.2, while controlling for the effects of the other mechanisms discussed in Section 2.1.

The experiment used a 4x2 between-subject design. In each of the eight sub-treatments, participants completed eleven tasks. These were made up of six *goods tasks*, where participants were presented with a good and a valuation was elicited for that good, and five cognitive distraction *lottery tasks*, which occurred between each pair of goods tasks.

3.1 Goods Tasks

The goods tasks involved the valuation of six different goods separated into two distinct subsets, three mugs and three chocolate bars. The use of two different subsets of goods allowed within-subset and between-subset novelty effects to be disentangled.

The three mugs were all white, ceramic mugs, and differed only in the type of pattern of shapes printed on the mug (‘Squares’, ‘Circles’ or ‘Triangles’). The three chocolate products were all the same luxury brand chocolate bar, differing only by the type of biscuit topping on each chocolate (‘Rocky Road’, ‘Milk and Cookies’ or ‘Mississippi Mud Pie’).[2] Thus, within each subset, the three goods differed only in subjective value. That is, they differed only on a dimension that was clearly a matter of personal taste, with no connotation of any difference in objective quality or market value. In real world individual consumption decision making it is common for goods compared for purchase to be identical in terms of price and quality, but to

[2] The three patterns printed on the mugs were custom-designed for the experiment. The three chocolate bars were from a luxury British confectioners (Hotel Chocolat). Detailed descriptions of the goods, which were shown to the participants when receiving the goods, can be found in Appendix A.

differ in terms of some dimension of personal taste, such as colour for items of clothing. This feature of the experiment was used to control for embedding effects that might occur if one good was perceived as objectively larger than or better than another, and to minimize the possibility that participants would use beliefs about market prices as cues for relative valuations within each subset. To the author's knowledge, there has been no research looking exclusively at the effect of order on valuations of private goods that differ only in terms of their subjective value.

The goods used in the experiment were determined in light of a pilot survey, with the objectives that: i) within each subset, there should be considerable cross-participant variation in preference rankings of the three goods; ii) for most individual participants, differences in valuations between goods in the same subset should be relatively small, iii) the types of goods used in the two subsets should be substantially different; iv) average valuations of goods in the two subsets should be similar to one another.

Once the experiment began, the procedure was identical across all treatments. Participants were told in each goods task that they would be given a good which they then conditionally owned, but that they would be able to sell back to the experimenter at a price that would be determined at the end of the experiment. A good was distributed to participants, and they were encouraged to look at (and pick up) the good to assess it, before making a series of binary decisions designed to elicit the minimum price at which participants were willing to sell that good.

Eliciting selling, rather than buying, valuations was intentional. The experiment aimed to manipulate dimensions of the attributes of the goods and their relative novelty value to affect attention levels. As noted in Section 2.2, previous research suggests that sellers give more attention than buyers to the attributes of goods (Carmon and Ariely, 2000, Nayakankuppam and Mishra, 2005). Thus, attention- and novelty-based effects are more likely to be found in selling tasks. It has previously been acknowledged that the growth of online second-hand markets has increased the frequency with which consumers take the role of sellers, as opposed to the more conventional role of buyers (Simonson and Drolet, 2004, pp.681-682), suggesting that such a position would not be an unfamiliar one to participants.

A common feature of experiments that exhibit shaping effects, or evidence of strategic bidding, is the occurrence of feedback about the outcome of each task (for example, about whether the participant has sold the good and if so, at what price) before the next task begins. In this experiment, there was no such feedback. A random lottery incentive system was used: one task (of the total eleven) was randomly selected, at the end of the experiment, to determine participants' earnings. This randomly selected task was common for all participants. This meant that the selected task would either be a goods task or a lottery task for all participants, but the counterbalancing of specific tasks within each task type meant that the specific good or lottery selected would differ across participants.

If the selected task was a goods task, the price at which the good could be sold would only then be revealed. Thus, there was no possibility for participants' responses to later tasks to be influenced by prices revealed in earlier tasks. Because valuations for each good were elicited immediately after the good was shown to participants, order effects could not be caused by differences in the strength of memory retrieval over time.

The downward trends in selling valuations found by Shogren *et al*. (1994) and Loomes *et al*. (2010) may be the result of a reduced feeling of ownership (and so reduced loss aversion) for later goods. In these studies, a market valuation was determined after each round. Participants knew the outcome of their decision in each round, and so may have 'sold' goods in multiple rounds. If participants were frequently 'selling' their goods after completing goods tasks, this might reduce loss aversion in later tasks. This effect cannot occur in the present design because of the absence of feedback. Since no selling decision is realised until the end of the experiment, a sense of conditional ownership is maintained for every good. Whilst it is possible that there remains a diminished loss aversion that stems from participants simply completing multiple goods tasks and thinking about selling, such an effect can be classified as an experimental novelty effect – a tendency for participants to give less attention to losses as the experiment progresses.

As noted in Section 2.1, order effects in contingent valuation studies have sometimes been attributed to the lower credibility of scenarios that appear later in an experiment or survey. Effects of this kind are unlikely in the present design, in which participants report valuations for real, private goods and in which the incentive system is clearly defined.

If participants fail to behave in accordance with expected utility theory, responses elicited using a random lottery incentive system may differ from those elicited in single-task designs. Existing evidence suggests that this does not induce systematic bias for simple choices between lotteries (Cubitt *et al.*, 1998), but that selling valuations tend to be lower in random-lottery designs than in single-task designs (Loewenstein and Adler, 1995). However, this potential bias is not a problem for the present study, which is concerned only with the relative valuations of goods across tasks.

In order to maximise the likelihood of honest valuation decisions of participants, a Becker-DeGroot-Marschak (henceforth BDM) mechanism (1964) was used. In each goods task, participants were shown a set of possible prices, ranging from £0.20-£6.00, in £0.20 increments, and were asked whether they wished to not sell (and so keep) the good, or to sell it back to the experimenter at each of the those prices. Once all participants had completed the valuation decisions for that task, all goods were collected and the experiment continued. This was kept the same for all goods tasks.

The £6.00 upper limit was chosen in the expectation that most participants would value the goods less than this. (The market values of the goods, which were not revealed to participants, were considerably less than £6.00.[3]) This design choice reduced the possibility that participants' valuations might be framed by the upper and lower bounds of the BDM mechanism (Bohm *et al*., 1997).

A participant who acted on consistent preferences between money and goods would report at most one *preference switch*, from 'not sell' at relatively low prices to 'sell' at relatively high prices. (There would be no preference switch for a participant who would 'not sell' at every price or would 'sell' at every price). If a participant's decisions implied more than one preference switch, these switches were highlighted on the participant's screen, together with the relevant material from the experimental instructions which reminded them of the workings of the valuation mechanism, and gave them the opportunity to revise their decision, if they wished to. Participants were free to resubmit decisions with two or more preference switches, but this revision stage allowed errors to be corrected.

If a participant reported no more than one preference switch for a good, the location of that switch (or its absence) locates one of thirty-one points on an ordinal valuation scale. For a participant with exactly one preference switch, their *valuation* of that good will be defined as the mean of the highest price at which they would 'not sell' and the lowest price at which they would 'sell' (or, equivalently, the lowest price at which they would 'sell', *minus* £0.10).

[3] The retail values of the goods were: mugs- £2.25 and chocolates- £3.15, although pre-experimental surveys suggested that participants' willingness-to-pay for the two were both approximately £1.50.

Participants who would ‘sell’ at every price will be defined to have a valuation of £0.10, and those who would ‘not sell’ at every price to have a valuation of £6.10.

3.2 Cognitive Distraction Tasks

In the five lottery tasks, participants were asked to choose which of two monetary lotteries they would prefer to play, with payoffs determined by the roll of a die. Lottery task (i) is shown as an example in Figure 1, in the form in which it was presented on participants’ computer screens. During the experiment, the possible outcomes of the lotteries were shown as fractions of *X*, with participants aware that *X* could take one of thirty values, from £0.20 to £6.00 in £0.20 increments. Notice that the set of possible values of *X* is the same as the set of possible prices in the goods tasks. This *X*-value lottery design ensures that the two types of tasks involve the same reference points, so that comparisons of goods valuations are not distorted by values used in the lottery tasks.

Whilst the primary function of the lottery tasks was to act as a cognitive distraction between goods tasks, the responses of participants to the lottery tasks revealed risk attitudes and possible violations of expected utility theory. The order in which participants completed the lottery tasks was counterbalanced across participants, and the position of the lotteries within each task was counterbalanced also. All lottery task descriptions can be found in Appendix B.

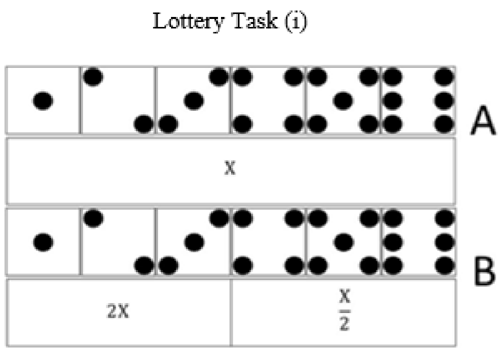


Figure 1. An example of a lottery task

3.3 Payments to Participants

Once all tasks were completed, one of the eleven task numbers was then drawn at random to determine the task for which each participant would play for real. After this, one of the thirty values (from £0.20 to £6.00) was drawn at random, and this was determined to be either the price of the good or the value of *X* in the lotteries.

If the task drawn for a participant was a goods task, the participant's decision about whether to sell the good at the drawn price was made binding. If the participant *was not* willing to sell the good at the drawn price, they kept the good and took it away with them (in addition to a £6 participation fee). If the participant *was* willing to sell the good at that price, they did not take the good away but instead received the drawn price in addition to their £6 participation fee. The fact that truthful valuations were the optimal response in the valuation mechanism was made clear to participants in both the instructions and the pre-experimental quizzes.

If the task drawn for a participant was a lottery task, the participant was then shown the lottery they had chosen in that task with the value of X equal to the drawn value. An experimenter visited them with a die to determine their final payoff, in addition to the £6 participation fee.

3.4 Treatments

The experiment had four treatments (A–D), each of which was subdivided into two sub-treatments (e.g. A1 and A2). Treatments A, B and C, which will be called the *main treatments*, differ only in the order in which goods in the goods tasks were presented. Treatment D uses the same order of goods as Treatment A, but adds a control designed to remove good-specific novelty from the experimental design. This control uses a method of *advanced disclosure* that has been used in some contingent valuation surveys, in which respondents are fully informed about the types of goods to be valued before any valuations are elicited (Bateman *et al*., 2004). In Treatment D, before facing any of the tasks, participants were shown on their screens images of the six goods to be valued, and were informed these would be the goods for valuation in the six goods tasks. The order in which these tasks would be presented was not revealed at this stage. (In fact the order was counterbalanced such that, for half the participants, the task order reproduced the top-to-bottom-left-to-right order in which the goods were initially presented; for the other half, this task order was reversed.)

In all treatments, the same three different chocolates and three different mugs were used as the goods for valuation. The order in which the goods were presented in each sub-treatment is shown in Table 1. M_1, M_2 and M_3 respectively denote the first, second and third mug task faced by a participant; C_1, C_2 and C_3 denote the first, second and third chocolate tasks. Within each treatment, which of the three mugs (Squares, Circles or Triangles) appeared in which of the positions M_1, M_2 and M_3 was counterbalanced, and similarly for the three chocolates. Thus, for example, Table 1 shows that in Sub-Treatment A1, participants faced three tasks involving mugs followed by three tasks involving chocolate. Notice that, within each treatment, the two sub-treatments counterbalance mugs and chocolates. Thus, for example, Treatment A can be interpreted as a treatment in which participants face three tasks involving one type of good followed by three tasks involving the other; whether mugs are faced first or second is counterbalanced.

The main treatments share the feature that there is either one transition between good types (Treatment A) or two transitions (Treatments B and C). All task orders that are compatible with this constraint are included in the design. By imposing this constraint, rather than counterbalancing all possible task orders, the design increases the power of tests of within-subset novelty. (For example, it ensures that for two-thirds of all main-treatment participants, the three mug tasks are faced in succession.)

Considering only the main treatments, each task can be described by four characteristics: *order* (i), *within-subset novelty* (j), *between-subset novelty* (k), and *good type* (m). Order takes the form $i \in \{1, \ldots, 6\}$, with i=1, …, 6 referring to the first, …, sixth goods task faced. Within-subset novelty takes the form $j \in \{1, 2, 3\}$, with j =1, 2, 3 referring to the whether the task is the first, second and third involving a good of the relevant type. Between-subset novelty takes the form $k \in \{1, 2\}$, with k=1 referring to a task in which the good is of the same type as the good faced in the first task, and k=2 to a case in which it is not. Good type takes the form $m \in \{0, 1\}$, where m=0 refers to a task involving a chocolate and m=1 refers to a task involving a mug.

		Task Order					
		1	2	3	4	5	6
Treatment	**A1**	M_1	M_2	M_3	C_1	C_2	C_3
	A2	C_1	C_2	C_3	M_1	M_2	M_3
	B1	M_1	M_2	C_1	C_2	C_3	M_3
	B2	C_1	C_2	M_1	M_2	M_3	C_3
	C1	M_1	C_1	C_2	C_3	M_2	M_3
	C2	C_1	M_1	M_2	M_3	C_2	C_3
	D1	M_1	M_2	M_3	C_1	C_2	C_3
	D2	C_1	C_2	C_3	M_1	M_2	M_3

Table 1. Order of goods tasks by sub-treatments

Thus, for example, the second task in Treatment A1 is described by i=2, j=2, k=1, m=1; the fifth task in Treatment C2 is described by i=5, j=2, k=1, m=0. As a matter of definition, these four characteristics are not completely independent of one another (for example, i=1 necessarily implies j=1 and k=1). However, the design ensures that the impact on valuations of variation in each characteristic can be captured in isolation, holding other characteristics constant.

3.5 Implementation

The experiment took place in early 2016 at the University of East Anglia's Centre for Behavioural and Experiment Social Science (CBESS). The 243 participants were recruited through the centre's online recruitment system and had no experience of experiments of this type. The experiment was conducted using experimental software package z-Tree (Zurich Toolbox for Ready-made Economic Experiments) (Fischbacher, 2007).

The experimental instructions were read aloud and participants had the opportunity to ask any questions. Before facing the goods and lottery tasks, participants were asked to answer a set of multiple-choice questions to test their understanding of the experimental procedure, including questions regarding the BDM valuation mechanism in the goods tasks and possible lottery outcomes in the lottery tasks. If a participant answered incorrectly on the first attempt, they were asked to review the relevant instructions and attempt the question again. 83.48% of questions were answered correctly at the first attempt, and 95.98% were answered correctly by the second attempt, suggesting participants in general understood the mechanisms of the experimental design.

4. Hypotheses

The hypotheses that refer to Treatments A, B and C can be formulated in terms of the distribution, within the population of potential participants, of valuations $V_{i,j,k,m}$, where $V_{i,j,k,m}$ denotes a valuation that is conditional on order *i*, within-subset novelty *j*, between-subset novelty *k*, and good type *m*. (Within either good type, the specific good to which the valuation refers is to be interpreted as a random draw from the relevant set of three goods.) The assumption that participants act on neoclassical preferences implies the null hypothesis that the distribution of $V_{i,j,k,m}$ is independent of the values of *i*, *j* and *k*. As explained in Sections 2.2 and 3.4, the experiment was designed to test the following hypotheses about attention-based order effects:

Alternative Hypothesis H1: Experimental novelty effects

(a) For all feasible j,k,m: $V_{1,j,k,m} > V_{2,j,k,m} > V_{3,j,k,m} > V_{4,j,k,m} > V_{5,j,k,m} > V_{6,j,k,m}$

(b) For all feasible j,k,m, and for all $s \in \{2, 3, 4, 5\}$: $(V_{1,j,k,m} - V_{2,j,k,m}) > (V_{s,j,k,m} - V_{(s+1),j,k,m})$

Part (a) of this hypothesis predicts that, holding all other factors constant, the later in the series of tasks a good is valued, the lower its valuation. Part (b) predicts a *primacy effect*: holding all other factors constant, experimental novelty effects are stronger between the first and second goods task than between other pairs of adjacent goods tasks. For example, consider Sub-Treatments A2 and B2. In B2, the first mug task, M_1, is the third goods task, whereas in A2, the first mug task, M_1, is the fourth goods task. Hypothesis H1(a) predicts that M_1 elicits a higher valuation in sub-treatment B2 than in A2.

Alternative Hypothesis H2: Within-subset novelty effects

For all feasible i,k,m: $V_{i,1,k,m} > V_{i,2,k,m} > V_{i,3,k,m}$

This hypothesis predicts that, holding all other factors constant, the less novel a good within a subset, i.e. the more goods of that subset that have already been valued, the lower its valuation. For example, consider Sub-Treatments A2 and C2. In A2, the fourth goods task is M_1, i.e. the first task to involve a mug. In C2, the fourth goods task is M_3, i.e. the third task to involve a mug. Hypothesis H2 predicts that the valuation elicited in the fourth goods task of A2 is greater than that elicited in the fourth goods task of C2.

Alternative Hypothesis H3: Between-subset novelty effects

For all feasible i,j,m: $V_{i,j,1,m} > V_{i,j,2,m}$

This hypothesis predicts that, holding all other factors constant, the first good type presented has a higher valuation than the second good type. For example, consider Sub-Treatments A2 and C1. In both sub-treatments, the fifth goods task is M_2, i.e. the second task to involve a mug. In A2, the first good type presented was mugs; in C1, it was chocolates. Hypothesis H3 predicts that the valuation elicited in the fifth goods task of A2 is greater than that elicited in C1.

Alternative Hypothesis H4: Advanced disclosure

Whilst this is not the primary objective of this paper, it is also of interest to consider how far advanced disclosure reduces attention-based order effects. Hypothesis H4 predicts that within-subset and between-subset novelty effects, as predicted by Hypotheses H2 and H3, are less strong in Treatment D than in the main treatments.

5. Results

A total of 243 participants took part in the experiment, but 11 participants reported inconsistent valuation decisions such that unambiguous intended valuations could not be inferred, and so were dropped from the analysis[4]. This left 232 participants in total with usable data.

5.1 Summary Statistics

It is first of interest to observe valuations across order by treatment. Table 2 reports the mean valuation in each of the four treatments, pooling across the two component sub-treatments. Figure 2 presents the same data graphically.

		Task Order						
Valuation (£)		1	2	3	4	5	6	*n*
Treatment	**A**	2.53	2.32	2.37	2.36	2.40	2.35	58
	B	2.60	2.56	1.96	1.93	2.06	2.38	60
	C	2.29	1.87	1.84	1.74	1.87	1.91	60
	Pooled A-C	2.47	2.25	2.05	2.01	2.11	2.21	178
	D	2.41	2.32	2.20	2.09	2.00	1.98	54

Table 2. Mean valuations by treatment and task order

Pooling Treatments A to C, there is an overall downward trend in valuations. A non-parametric test (Cuzick, 1985) shows that this trend is statistically significant (z= 2.05, p=0.040). This is an extension of a Wilcoxon rank-sum test, testing for a consistent trend in differences in the value of a variable across multiple sets of observations, where these sets have a natural ordinal ranking. Treatment D shows a similar trend (z= 1.75, p=0.080).

[4]Recall participants were given an opportunity to rectify apparently inconsistent decisions in each goods task. If participants still revealed inconsistency after this, efforts were made to allow for human error and still record an intended valuation. This was achieved through the following rule: if consistency could be achieved through the rectification of *one* valuation decision, and it was obvious which valuation decision was erroneous, then this one valuation decision was rectified and valuation was inferred from these consistent valuation decisions. The valuations of two participants (each with inconsistent valuation decisions for two of six goods) were amended using this rule. The 11 participants dropped from the analysis either had at least one goods task which required more than one valuation decision to become consistent, or it was not obvious which valuation decision was erroneous.

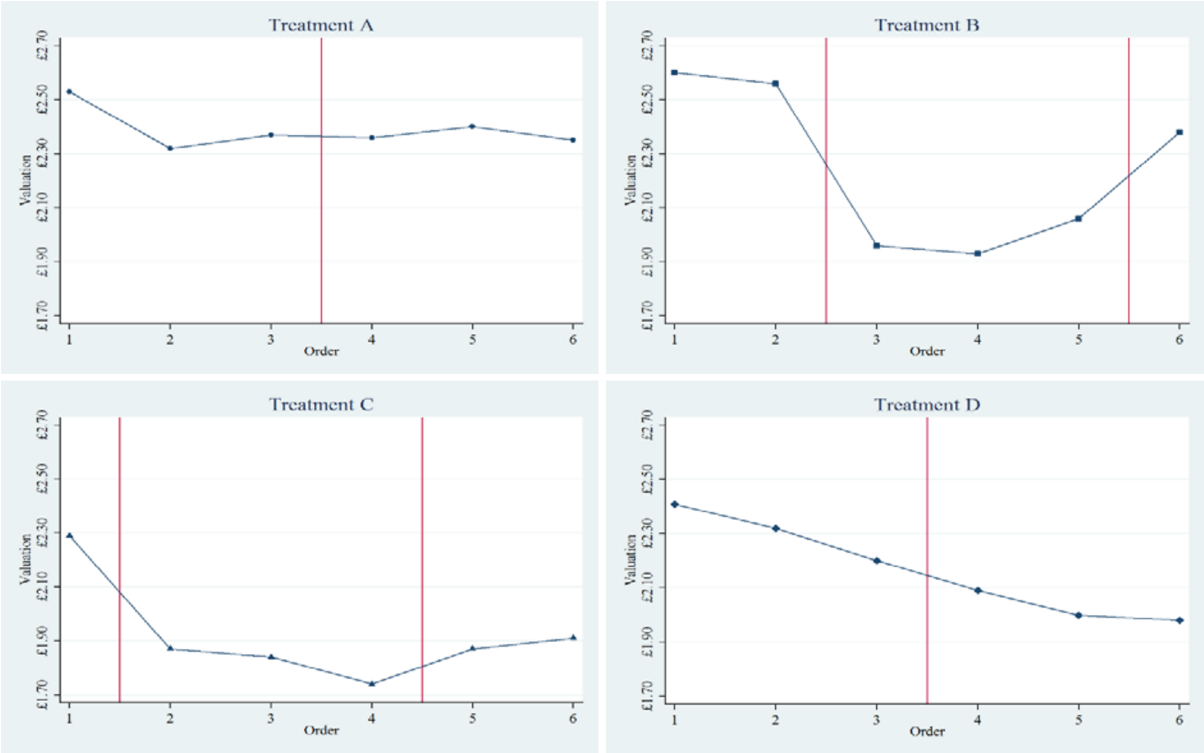


Figure 2. Graphs of valuations by treatment and task order

Notes: Vertical bars indicate a transition from one good type (m), to another.

5.2 Results: Tests of Hypotheses H1-H3

Whilst non-parametric analysis enabled a test for overall trends in valuations, parametric regression analysis is required to disentangle the different novelty effects predicted by Hypotheses H1–H3. Regression analysis also allows for the effects on valuations of demographic variables and of responses in lottery tasks to be included (regression results including these variables can be found in Appendix D).

Whilst valuations in this experiment were restricted to not be less than £0.10 or greater than £6.10, actual valuations could be less than greater than these limit valuations (though only 6 participants chose to not sell at least one good at £6.00). To address this econometrically, a Tobit model will be used, with lower and upper bounds set at £0.10 and £6.10 respectively.

Given the potential for the non-independence of valuations at the participant level, a random effects Tobit regression model is used. Table 3 reports the results of three estimated models pooling Treatments A-C. In each model, the dependent variable is the valuation $V_{i,j,k,m}$ reported by a participant in a task with order i, within-subset novelty j, between-subset novelty k and good type m. The following independent variables are used:

Order: takes the value 0 – i (i.e. 0, 1, …, 5 for tasks that appear in order 1, 2, …, 6).

First Task: takes the value 1 when i=1 (i.e. when the task is the first to be faced by the participant), 0 otherwise.

Novelty$_2$: takes the value 1 when j=2 (i.e. when the task is the second to involve a good of the relevant type), 0 otherwise.

Novelty$_3$: takes the value 1 when j=3, 0 otherwise (i.e. when the task is the third to involve a good of the relevant type), 0 otherwise.

First Good Type: takes the value 1 when k=1 (i.e. when the task involves the first good to be seen by the participant), 0 when k=2.

Mug: takes the value m (i.e. 1 if the good type is 'mug', 0 if it is 'chocolate').

In all three models, valuations are estimated to be £0.47 higher for chocolates than for mugs, and this difference is strongly significant (p< 0.001)

Valuation	(Model 1) **Hypothesis H1(a)**	(Model 2) **Hypothesis H1(a)** **Hypothesis H1(b)**	(Model 3) **Hypothesis H1(a)** **Hypothesis H1(b)** **Hypothesis H2** **Hypothesis H3**
Order	-0.0614***	-0.0090	-0.0083
	(0.016)	*(0.021)*	*(0.026)*
First Task	-------	0.3646***	0.2229*
	-------	*(0.094)*	*(0.119)*
Novelty$_2$	-------	-------	-0.0061
	-------	-------	*(0.088)*
Novelty$_3$	-------	-------	-0.0127
	-------	-------	*(0.104)*
First Good Type	-------	-------	0.2291***
	-------	-------	*(0.066)*
Mug	-0.4712***	-0.4706***	-0.4747***
	(0.054)	*(0.053)*	*(0.053)*
Constant	2.5069***	2.3151***	2.2301***
	(0.112)	*(0.122)*	*(0.131)*
Sigma(u)	1.3048***	1.3058***	1.3080***
	(0.076)	*(0.076)*	*(0.076)*
Sigma(e)	0.8614***	0.8535***	0.8469***
	(0.022)	*(0.022)*	*(0.021)*
# Obs	1,068	1,068	1,068
# Groups	178	178	178

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3. Random effects Tobit models of order, primacy, *between*- and *within-subset* novelty effects on valuation in treatments A-C (lower limit: £0.10, upper limit: £6.10)

Model 1 tests for overall order effects (i.e. general trends in valuations). In line with the non-parametric test reported in Section 5.1, there is strongly significant evidence of an order effect (p< 0.001). Valuations are estimated to fall by £0.06 in each successive task.

Model 2 drops the assumption that order effects are linear, by including the variable *First Task* which picks up primacy effects. This variable has a strongly significant positive effect on valuations (p< 0.001), but its inclusion makes *Order* insignificant.

Model 3 includes variables capable of separately identifying the attention-based effects predicted by Hypotheses H1–H3. After controlling for other effects, *First Task* remains positive and significant (p= 0.060); valuations in the first task are estimated to enjoy a premium of £0.22. *Order* remains insignificant; its estimated effect on valuations is virtually zero. $Novelty_2$ and $Novelty_3$, which would pick up the within-subset novelty effects predicted by Hypothesis H2, are insignificant and have estimated values close to zero. *First Good Type*, which picks up the between-subset novelty effects predicted by H3, is positive and strongly significant (p= 0.001). Valuations of the first good type to be faced are estimated to enjoy a premium of £0.23.

This analysis supports the following conclusions:

Result 1 (experimental novelty): Holding other factors constant, there is strong evidence of a primacy effect: valuations are higher in the first task than in all subsequent tasks. There is no evidence of a decline in valuations after the second task.

Result 2 (within-subset novelty): There is no evidence of the within-subset novelty effects predicted by Hypothesis H2.

Result 3 (between-subset novelty): There is strong evidence of the between-subset novelty effect predicted by Hypothesis H3.

5.3 Results: Tests of Hypothesis H4

Models 1, 2 and 3 were also estimated using only data from Treatment D, the advanced disclosure treatment. Table 4 reports the results of these estimations. Whilst order, primacy, between-subset novelty and good type variables are defined as before, $Novelty_2$ and $Novelty_3$ are pooled as $Novelty_{23}$ (i.e. takes the value 1 when j=2 or j=3, 0 otherwise) to avoid over-identification[5]. In comparing these results with those in Table 3, it should be noticed that the sample size is much smaller (54 rather than 178) and that statistical tests are correspondingly less powerful.[6]

Valuation	(Model 1) **Hypothesis H1(a)**	(Model 2) **Hypothesis H1(a)** **Hypothesis H1(b)**	(Model 3) **Hypothesis H1(a)** **Hypothesis H1(b)** **Hypothesis H2** **Hypothesis H3**
Order	-0.1056***	-0.1008***	-0.0662
	(0.024)	*(0.032)*	*(0.100)*
First Task	-------	0.0337	0.0605
	-------	*(0.145)*	*(0.174)*
Novelty3	-------	-------	-0.0099
	-------	-------	*(0.194)*
First Good Type	-------	-------	0.1153
	-------	-------	*(0.317)*
Mug	-0.4417***	-0.4415***	-0.4429***
	(0.082)	*(0.082)*	*(0.082)*
Constant	2.6176***	2.5998***	2.4584***
	(0.193)	*(0.208)*	*(0.363)*
Sigma(u)	1.2704***	1.2705***	1.2705***
	(0.130)	*(0.130)*	*(0.130)*
Sigma(e)	0.7289***	0.7289***	0.7285***
	(0.033)	*(0.033)*	*(0.033)*
# Obs	324	324	324
# Groups	54	54	54

Table 4. Random effects Tobit models of order, primacy, *between*- and *within-subset* novelty effects on valuation in treatment D (lower limit: £0.10, upper limit: £6.10)

Model 1 shows a strong and highly significant overall order effect ($p < 0.001$); valuations are estimated to fall by £0.11 in each successive task. As one would expect from a glance at the graph for Treatment D in Figure 2, Model 2 shows no significant primacy effect, and the estimated premium for the first task is very small (£0.03). The overall order effect remains strong and significant ($p = 0.002$). In Model 3, none of the order or novelty variables are significant, perhaps reflecting the small sample size and the high degree of positive correlation between the variables. Given the absence of any obvious discontinuities in the downward trend

[5] The problem of over-identification arises because, within Treatment D, there is insufficient counterbalancing of order and novelty variables.
[6] The two orders in which goods were presented at the advanced disclosure stage were counterbalanced across participants (see Section 3.4). The regressions reported in Table 4 pool these sub-treatments. There were no systematic differences in valuations between them (see Appendix C).

in the Figure 2 graph, it is natural to draw the following conclusion, which gives limited support for Hypothesis 4:

Result 4 (advanced disclosure). Under advanced disclosure, there is no evidence of good-specific (within-subset or between-subset) novelty effects.

5.4 Other Findings

Whilst the primary purpose of the lottery tasks was to act as a cognitive distraction between goods tasks, they elicited some information about participants' degrees of risk aversion and their propensities to violate principles of expected utility theory (EUT). Summary data about responses to these tasks can be found in Appendix B. A large majority of participants revealed risk aversion. 53.0 per cent of participants revealed the common ratio effect, one of the most commonly-observed violations of the independence axiom of EUT (see, for example, Cubitt *et al*., 1998), while only 4.7 per cent violated that axiom in the opposite direction, a discrepancy consistent with other experiments (e.g. Starmer and Sugden, 1989). There were also very few violations of the dominance axiom, and these observations suggest that participants understood the *X*-value lottery design. Participants' valuations in the goods tasks were not significantly affected by their degree of risk aversion or by their propensities to violate EUT (see Appendix D).

Demographic information, collected in a post-experimental questionnaire, was included in additional regression analyses to test if any demographic factors consistently influenced participants' valuations. These questionnaires were optional. Six participants did not answer all the questions and so were omitted from the analysis, reported in Appendix D. Gender, nationality and previous formal study of economics had no significant effects, but there is some evidence that older participants reported lower valuations. The most interesting finding was that, in the main treatments, experience of having taken part in previous economics experiments had a strong and significant negative effect on valuations (p= 0.008).[7] Valuations were estimated to be £0.84 higher for participants who were taking part in an experiment for the first time. Adding variables that interacted experience with order and novelty revealed no obvious patterns. Viewed in the perspective of an attention-based theory, this effect of experience may suggest that, the more frequently participants take part in any experiment, the less novel future experiments become, and so the attention participants give to the nature of them diminishes, reducing valuations in general.

[7] In Treatment D, for which the sample size was much smaller, no significant effects were found.

Because of the counterbalancing of the order of tasks (see Section 3.4), any systematic differences between the valuations of the three goods in each subset would not affect the tests of Hypotheses H1–H4. Nevertheless, it is of interest to assess whether there were any such differences. Two relevant tests are reported in Appendix E. These tests use data only from those participants who implicitly reported a 'most preferred' good in a subset (i.e. who gave one good a strictly higher valuation than each of the other two). The first test is of whether the distribution of first preferences is non-random across goods. The second test is of whether the absolute valuations of participants' most-preferred goods differs according to which good is most preferred. No significant differences are found for either test, applied to either subset.

6. Discussion

The results reported in Section 5 provide evidence that, when experimental participants face a series of tasks designed to elicit selling valuations *for different goods*, valuations tend to fall over the course of the experiment. This evidence parallels previous experimental findings about the effects of repeating *the same* selling task.

By using goods with different degrees of similarity to one another in different orders, the experiment was able to disentangle different attention-based mechanisms that might induce a fall in valuations. The treatments without advanced disclosure produce three main results. First, there is a strong tendency for participants' valuations to be higher in the first task they face than in subsequent tasks – an *experimental* novelty effect. Second, in a setting in which goods fall into two dissimilar types, there is a strong tendency for valuations to be higher for the first type of good faced than for the second – a *between-subset* novelty effect. Third, there is no evidence of *within-subset* novelty effects.

This combination of results may seem surprising. Given the known tendency for valuations of identical goods to decline as tasks are repeated, one might have expected to find the same tendency when goods are very similar to one another. The absence of a within-subset novelty effect suggests that apparently small differences between goods – slightly different patterns on otherwise identical mugs, different biscuit toppings on otherwise identical chocolate bars – can maintain participants' attention in a sequence of goods tasks. It may be significant that, for each good type, these 'small' differences were restricted to one attribute of the good. The fact that this was the *only* attribute that varied between the relevant tasks would have made it particularly salient to participants. A participant whose attention is focused on biscuit toppings on chocolate bars (and who is anticipating the possibility of having one to eat) can experience a sequence of ‘Rocky Road’, ‘Milk and Cookies’ and ‘Mississippi Mud Pie’ as three distinct novelties.

In contrast, the between-subset novelty effect implies that the transition from valuing mugs to valuing chocolate bars (or vice versa) induced a reduction in attention. It seems that some of the novelty of the first task carried over to later tasks in which the same type of good appeared, but not to tasks involving a different type. A possible explanation of this effect is that it is a form of anchoring, analogous with the finding of Payne *et al*. (2000) that when respondents sequentially report valuations for each of a given set of public projects, the sum of these valuations is influenced by the order in which valuations are made: the higher the *relative* valuation of the good faced first, the higher is the sum of valuations.

This effect might work through anchoring *on valuations*: participants might use valuations that they have reported in earlier tasks as anchors when subsequently valuing similar goods. But, given that the effect does not apply across good types (while, as noted in Section 2.1, even arbitrary numbers can be anchors for valuations) it is more plausible to conjecture that anchoring is *on attributes*. As explained in Section 2.2, theories of attentional bias explain differences between buying and selling valuations as a result of sellers giving more attention than buyers to the (positive) attributes of goods. Relatedly, it is possible that the positive attention devoted to the attributes of the first valued good then acts as an anchor for subsequent valuations. Participants may attend more to the attributes that are possessed by that good than to attributes that they experience later. Thus, the initial focus of attention of these attributes may spill over to different goods of the same type – that is, goods that share many of the attributes of the first good.

These conjectures receive some support from the results of the advanced disclosure treatment. The distinguishing feature of this treatment was that participants saw all six goods before facing the first goods task: individual goods tasks were not associated with the novelty of learning about a new good. In this treatment, unlike the other treatments but in line with previous experiments using identical goods, there was a consistent decline in valuations over the six tasks. It may be that advanced knowledge of all the goods made later tasks less interesting, reducing the attention that participants gave to them, independent of the good to be valued in any particular task. One must be careful not to over-interpret the results of this treatment, given its small sample size, but the absence of significant primacy and between-subset novelty effects in this treatment is consistent with the conjecture that advanced disclosure dampened participants’ sense of novelty when facing successive goods tasks.

7. Conclusion

In designing and interpreting experiments and stated preference methodologies that elicit individuals’ valuations of goods, it is important to understand the mechanisms by which participants’ responses to tasks can be affected by the order in which those tasks are faced. The experiment reported in this paper is a contribution to this under-researched area. Its findings highlight the importance of attention in mediating order effects, and the potentially complex relationships between the novelty of a task and the attention that it receives. In economics experiments, the quality of the data typically depends on participants’ engagement with and attention to the tasks they face. A fuller understanding of the role of novelty in maintaining attention may lead to more effective experimental designs.

Acknowledgements

The author would like to give thanks to Robert Sugden and Anders Poulsen for their guidance, comments and suggestions, and to the Centre for Behavioural and Experimental Social Science (CBESS) and the University of East Anglia for financial support. Thanks are also due to participants at various seminars at University of East Anglia and University of Amsterdam for helpful comments and feedback.

References

Ariely, D., Loewenstein, G. and Prelec, D., (2003), “Coherent Arbitrariness”: Stable Demand Curves without Stable Preferences. *Quarterly Journal of Economics*, *118*, 73-106

Bateman, I., Kahneman, D., Munro, A., Starmer, C. and Sugden, R. (2005). Testing Competing Models of Loss Aversion: An Adversarial Collaboration, *Journal of Public Economics*, *89*, 1561-1580.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997), Does Part–Whole Bias Exist? An Experimental Investigation, *Economic Journal*, *107*, 322-332

Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G.L. (2004), On Visible Choice Sets and Scope Sensitivity, *Journal of Environmental Economics and Management*, *47*, 71-93

Becker, G.M., DeGroot, M.H., and Marschak, J. (1964), Measuring Utility by a Single-Response Sequential Method, *Behavioral Science*, *9*, 226-232

Berlyne, D.E. (1951), Attention to Change, *British Journal of Psychology*, *42*, 269-278

Betsch, T., Fiedler, K., and Brinkman J. (1998), Behavioral Routines in Decision Making: The Effects of Novelty in Task Presentation and Time Pressure on Routine Maintenance and Deviation, *European Journal of Social Psychology*, *28*, 861-878

Bohm, P., Linden, J., and Sonnegard, J. (1997), Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets, *Economic Journal*, *107*, 1079-1089

Bruce, D., and Papay, J.P. (1970), Primacy Effects in Single-Trial Free Recall, *Journal of Verbal Learning and Verbal Behavior*, *9*, 473-486

Carmon, Z., and Ariely, D. (2000), Focusing on the Forgone: How Value Can Appear So Different to Buyers and Sellers, *Journal of Consumer Research*, *27*, 360-370

Carson, R.T. and Groves, T. (2007), Incentive and Informational Properties of Preference Questions, *Environmental and Resource Economics*, *37*, 181-210

Clark, J., and Friesen, L. (2008), The Causes of Order Effects in Contingent Valuation Surveys: An Experimental Investigation, *Journal of Environmental Economics and Management*, *56*, 195-206

Cubitt, R.P., Starmer, C. and Sugden, R. (1998), On the Validity of the Random Lottery Incentive System, *Experimental Economics*, *1*, 115-131

Cuzick, J. (1985), A Wilcoxon-Type Test for Trend, *Statistics in Medicine*, *4*, 543-547

Day, B., and Pinto Prades, J.L. (2010), Ordering Anomalies in Choice Experiments, *Journal of Environmental Economics and Management*, *59*, 271-285

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P. (2012), Ordering Effects and Choice Set Awareness in Repeat-Response Stated Preference Studies, *Journal of Environmental Economics and Management*, *63*, 73-91

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, *Experimental Economics*, *10*, 171-178

Herriges, J.A. and Shogren, J.F. (1996), Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning, *Journal of Environmental Economics and Management*, *30*, 112-131

Hogarth, R.M. and Einhorn, H.J. (1992), Order Effects in Belief Updating: The Belief-Adjustment Model, *Cognitive Psychology*, *24*, 1-55

Jaeger, S.R., and Harker, F.R. (2005), Consumer Evaluation of Novel Kiwifruit: Willingness-to-Pay, *Journal of the Science of Food and Agriculture*, *85*, 2519-2526

Kahneman, D., and Tversky, A. (1979), Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, *47*, 263-292

Kahneman, D., Knetsch, J.L., and Thaler, R. (1990), Experimental Tests of the Endowment Effect and the Coase Theorem, *Journal of Political Economy*, *98*, 1325-1348

Kahneman, D. and Knetsch, J.L. (1992), Valuing Public Goods: The Purchase of Moral Satisfaction, *Journal of Environmental Economics and Management*, *22*, 57-70

Knetsch, J.L., Tang, F.F. and Thaler, R.H. (2001), The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?, *Experimental Economics*, *4*, 257-269

Loewenstein, G., and Adler, D. (1995), A Bias in the Prediction of Tastes, *Economic Journal*, *105*, 929-937

Loomes, G., Starmer, C. and Sugden, R. (2003), Do Anomalies Disappear in Repeated Markets?, *Economic Journal*, *113*, 153-166

Loomes, G., Starmer, C., and Sugden, R. (2010), Preference Reversals and Disparities between Willingness-to-Pay and Willingness-to-Accept in Repeated Markets, *Journal of Economic Psychology*, *31*, 374-387

Meenakshi, J., Banerji, A., Manyong, V., Tomlins, K., Mittal, N., and Hamukwala, P. (2012), Using a Discrete Choice Experiment to Elicit the Demand for a Nutritious Food: Willingness-to-Pay for Orange Maize in Rural Zambia, *Journal of Health Economics*, *31*, 62-71

Nayakankuppam, D., and Mishra, H. (2005), The Endowment Effect: Rose-Tinted and Dark-Tinted Glasses, *Journal of Consumer Research*, *32*, 390-395

Pandelaere, M., Millet, K., and Van den Bergh, B. (2010), Madonna or Don McLean? The Effect of Order of Exposure on Relative Liking, *Journal of Consumer Psychology*, *20*, 442-451

Payne, J., Schkade, D., Desvousges, W., and Aultman, C. (2000), Valuation of Multiple Environmental Programs, *Journal of Risk and Uncertainty*, *21*, 95-115

Savage, S. J., and Waldman, D.M. (2008), Learning and Fatigue during Choice Experiments: A Comparison of Online and Mail Survey Modes, *Journal of Applied Econometrics*, *23*, 351-371

Scitovsky, T. (1992), *The Joyless Economy: The Psychology of Human Satisfaction*, Second Revised Edition, Oxford University Press

Shogren, J.F., Shin, S.Y., Hayes, D.J. and Kliebenstein, J.B., (1994), Resolving Differences in Willingness to Pay and Willingness to Accept, *American Economic Review*, 255-270

Shogren, J., Cho, S., Koo, C., List, J., Park, C., Polo, P. and Wilhelmi, R. (2001), Auction Mechanisms and the Measurement of WTP and WTA, *Resource and Energy Economics*, *23*, 97–109

Simonson, I., and Drolet, A. (2004), Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept, *Journal of Consumer Research*, *31*, 681-690

Starmer, C. and Sugden, R. (1989), Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect, *Journal of Risk and Uncertainty*, *2*, 159-178

Stevens, R., and Winter-Nelson, A. (2008), Consumer Acceptance of Pro-Vitamin A-Bio-Fortified Maize in Maputo, Mozambique, *Food Policy, 33*, 341-351

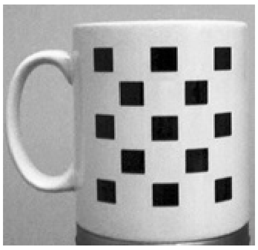
Sugden, R., Zheng, J., and Zizzo, D.J. (2013), Not All Anchors Are Created Equal, *Journal of Economic Psychology*, 39, 21-31

Tom, G. (2004), The Endowment-Institutional Affinity Effect, *Journal of Psychology*, *138*, 160-170

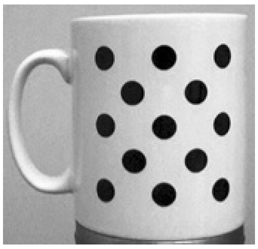
Tom, G., Nelson, C., Srzentic, T., and King R. (2007), Mere Exposure and the Endowment Effect on Consumer Decision Making, *Journal of Psychology: Interdisciplinary and Applied*, *141*, 117-125

Tversky, A. and Kahneman, D. (1974), Judgment under Uncertainty: Heuristics and Biases, *Science*, *185*, 1124-1131

Appendix A- Descriptions of goods used



A plain white ceramic mug with printed black squares. Dishwasher and microwave safe.



A plain white ceramic mug with printed black circles. Dishwasher and microwave safe.

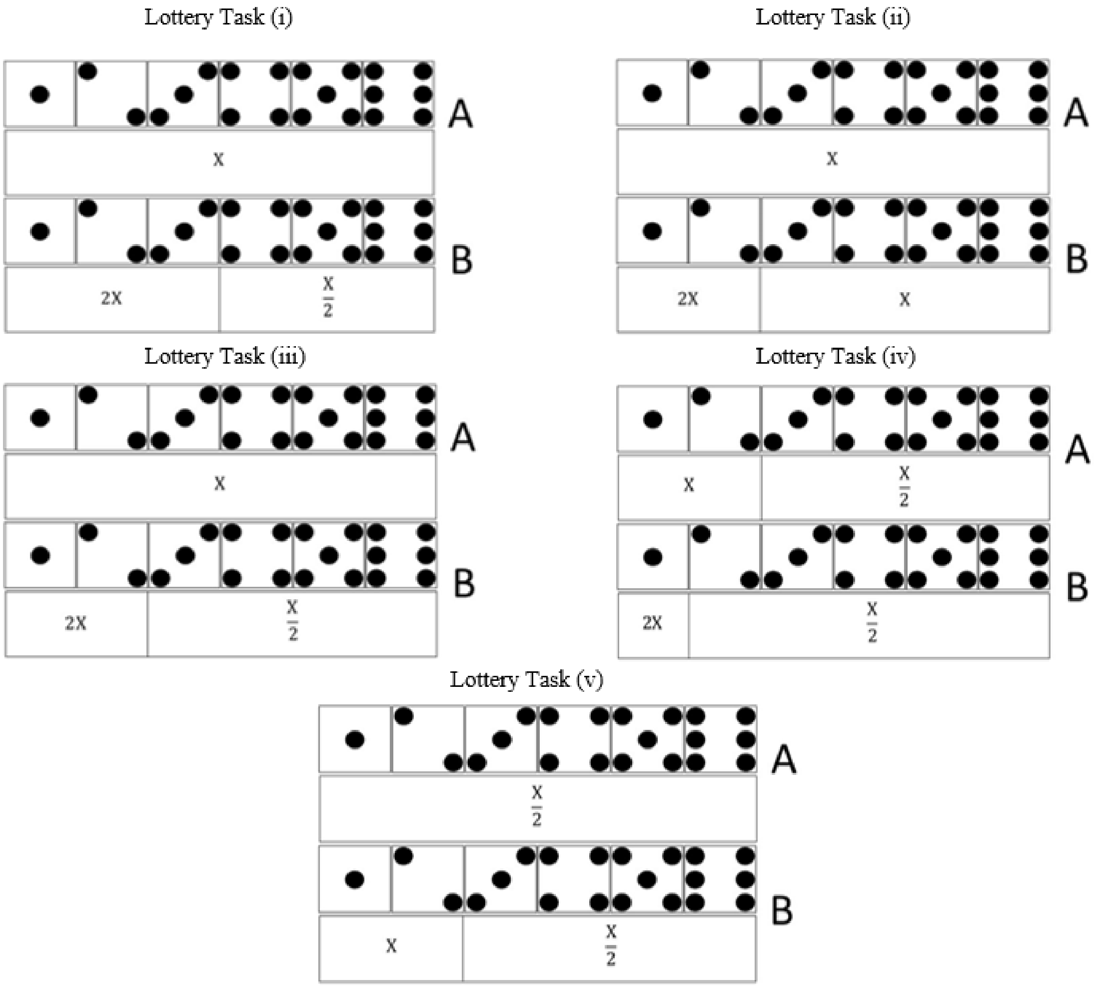
A plain white ceramic mug with printed black triangles. Dishwasher and microwave safe.

Rocky Road: Cookies and puffed rice set in milk and white chocolate.

Milk and Cookies: Shortcake and cocoa biscuits set in milk and white chocolate.

Mississippi Mud Pie: Cocoa biscuit crunch set in milk and white chocolate.

Appendix B- Lottery task descriptions and results



Risk Aversion: Lottery task (i)

Both options have the same expected payoff, but B involves risk and A does not. Thus, a risk averse subject would choose A, and a risk loving subject would choose B. 77.6% of participants revealed risk aversion.

Strong Risk Aversion: Lottery task (iii)

As the expected payoff of the risky lottery B exceeds the certain value of A, any subject choosing A could be seen as strongly risk averse. 28.4% of participants revealed strong risk aversion.

Dominance: Lottery task (ii) and lottery task (v)

In both tasks, Lottery B weakly dominates lottery A; thus a choice of A would violate the principle that preferences over lotteries respect stochastic dominance. In lottery task (ii), 5.6% of participants violated dominance, and lottery task (v), simply a scaled down version of (ii), 2.2% violated dominance. 0.09% of participants violated dominance in both lottery tasks.

Implied Dominance: Lottery task (i) and lottery task (iii)

Any subject choosing B in lottery task (i), but A in lottery task (iii) would indirectly violate dominance. 1.7% of participants indirectly violated dominance.

Independence Axiom- Common Ratio Effect: Lottery task (i) and lottery task (iv)

As lottery task (iv) is simply a scaled down version of lottery task (i), the axiom of independence implies that choices between A and B should be consistent across the two tasks. The choice of A in task (i) and B in task (iv) would represent the common ratio effect (Cubitt *et al*. (1998), for example). 53.0% of participants revealed the common ratio effect. 4.7% violated independence in the opposite direction.

Appendix C- Random effects Tobit model of order, primacy, *between*- and *within-subset* novelty effects on valuation by treatment *D* (lower limit: £0.10, upper limit: £6.10), including interaction variables for all effects on reverse presentation of advanced disclosure screen.

Valuation	(Model 1) **Hypothesis H1(a)**	(Model 2) **Hypothesis H1(a)** **Hypothesis H1(b)**	(Model 3) **Hypothesis H1(a)** **Hypothesis H1(b)** **Hypothesis H2** **Hypothesis H3**
Order	-0.1257***	-0.1155***	-0.0878
	(0.033)	*(0.044)*	*(0.139)*
First Task	-------	0.0703	0.1065
	-------	*(0.201)*	*(0.240)*
$Novelty_{23}$	-------	-------	0.0105
	-------	-------	*(0.269)*
First Good Type	-------	-------	0.0925
	-------	-------	*(0.440)*
Mug	-0.3687***	-0.3679***	-0.3708***
	(0.114)	*(0.114)*	*(0.114)*
Reverse AD	-0.0747	-0.0342	-0.0511
	(0.386)	*(0.415)*	*(0.726)*
Reverse AD*Order	0.0448	0.0339	0.0454
	(0.048)	*(0.064)*	*(0.200)*
Reverse AD*First Task	-------	-0.0757	-0.0985
	-------	*(0.290)*	*(0.347)*
Reverse AD*$Novelty_{23}$	-------	-------	-0.0429
	-------	-------	*(0.388)*
Reverse AD*First Good Type	-------	-------	0.0380
	-------	-------	*(0.633)*
Reverse AD*Mug	-0.1609	-0.1617	-0.1588
	(0.164)	*(0.164)*	*(0.164)*
Constant	2.6514***	2.6138***	2.4868***
	(0.269)	*(0.289)*	*(0.503)*
Sigma(u)	1.2715***	1.2717***	1.2717***
	(0.130)	*(0.130)*	*(0.130)*
Sigma(e)	0.7268***	0.7267***	0.7264***
	(0.033)	*(0.033)*	*(0.033)*
# Obs	324	324	324
# Groups	54	54	54

Appendix D- Random effects Tobit model of order, primacy, *between*- and *within-subset* novelty effects on valuation by treatments *A-C* and *D* (lower limit: £0.10, upper limit: £6.10), including lottery task and demographic effects, including interaction variables for all effects on non-experienced participants

Valuation	Lottery/ Demographics	Lottery/ Demographics	Lottery/ Demographics Non-Experience	Lottery/ Demographics Non-Experience
Treatments	*A-C*	*D*	*A-C*	*D*
Order	-0.0084	-0.0570	-0.0104	-0.0625
	(0.027)	*(0.101)*	*(0.028)*	*(0.100)*
First Task	0.2284*	0.0300	0.2287*	0.0575
	(0.122)	*(0.174)*	*(0.128)*	*(0.173)*
$Novelty_2$	-0.0088	-------	-0.0007	-------
	(0.090)	-------	*(0.095)*	-------
$Novelty_3$	-0.0162	-------	-0.0282	-------
	(0.106)	-------	*(0.112)*	-------
$Novelty_{23}$	-------	-0.0296	-------	-0.0287
	-------	*(0.195)*	-------	*(0.194)*
First Good Type	0.2284***	0.1266	0.2560***	0.2082
	(0.067)	*(0.318)*	*(0.071)*	*(0.316)*
Mug	-0.4753***	-0.4682***	-0.4685***	-0.4753***
	(0.054)	*(0.082)*	*(0.057)*	*(0.082)*
Risk Averse	-0.0106	-0.2120	-0.0096	-0.2103
	(0.318)	*(0.549)*	*(0.318)*	*(0.547)*
Strong Risk Averse	0.1806	-0.0680	0.1797	-0.0681
	(0.232)	*(0.440)*	*(0.232)*	*(0.440)*
Dominance	0.5839	-0.3960	0.5820	-0.3971
	(0.356)	*(0.939)*	*(0.356)*	*(0.937)*
Implied Dominance	-0.0114	1.1197	-0.0105	1.1187
	(0.801)	*(1.404)*	*(0.801)*	*(1.401)*
Common Ratio Effect	-0.3318	-0.0946	-0.3321	-0.0976
	(0.230)	*(0.466)*	*(0.230)*	*(0.465)*
Opposite Common Ratio Effect	0.0603	-0.0646	0.0608	-0.0660
	(0.516)	*(0.872)*	*(0.516)*	*(0.870)*
Age	0.0997	-0.2530	0.0996	-0.2527
	(0.079)	*(0.219)*	*(0.079)*	*(0.219)*
Age^2	-0.0140**	0.0180	-0.0140**	0.0180
	(0.006)	*(0.022)*	*(0.006)*	*(0.022)*
Female	-0.1254	0.1993	-0.1255	0.2010
	(0.204)	*(0.367)*	*(0.204)*	*(0.366)*
Economics	0.2354	-0.6167	0.2353	-0.6175
	(0.263)	*(0.445)*	*(0.263)*	*(0.444)*
UK/ Irish Nationality	-0.2773	-0.1813	-0.2770	-0.1791
	(0.218)	*(0.453)*	*(0.218)*	*(0.452)*
Non-Experience	0.8358***	0.2757	1.0247**	0.6748
	(0.317)	*(0.674)*	*(0.467)*	*(1.312)*
Non-Experience*Order	-------	-------	0.0137	0.0625
	-------	-------	*(0.110)*	*(0.356)*
Non-Experience*First Task	-------	-------	-0.0123	-0.3575
	-------	-------	*(0.429)*	*(0.617)*
Non-Experience*$Novelty_2$	-------	-------	-0.0744	-------
	-------	-------	*(0.294)*	-------
Non-Experience*$Novelty_3$	-------	-------	0.1195	-------
	-------	-------	*(0.385)*	-------
Non-Experience*$Novelty_{23}$	-------	-------	-------	0.0037
	-------	-------	-------	*(0.690)*
Non-Experience*First Good Type	-------	-------	-0.3208	-1.0582
	-------	-------	*(0.224)*	*(1.126)*
Non-Experience*Mug	-------	-------	-0.1556	0.0586
	-------	-------	*(0.211)*	*(0.291)*
Constant	2.3993***	3.1788***	2.3880***	3.1492***
	(0.369)	*(0.761)*	*(0.370)*	*(0.759)*
Sigma(u)	1.1952***	1.1626***	1.1953***	1.1635***
	(0.071)	*(0.122)*	*(0.071)*	*(0.121)*
Sigma(e)	0.8568***	0.7160***	0.8543***	0.6836***
	(0.022)	*(0.033)*	*(0.022)*	*(0.031)*
# Obs	1,044	312	1,044	312
# Groups	174	52	174	52

Appendix E- Tests for systematic preferences or valuations of specific goods

This table reports the distribution of good type preferences of participants who valued one good within a subset uniquely more highly than the other two (thus excluding participants who value more than one good equally most highly). The table pools all treatments *A*-*D* and also separates preferences by first subset of goods.

Mugs	**(n)**	**Squares**	**Circles**	**Triangles**	**χ^2 test**	***p*-value**
All	109	32	38	39	$\chi^2(2)= 0.790$	0.674
Mug First	68	21	24	23	$\chi^2(2)= 0.210$	0.902
Chocolate First	41	11	14	16	$\chi^2(2)= 0.930$	0.629
Chocolates	**(n)**	**RR**	**MC**	**MM**	**χ^2 test**	***p*-value**
All	115	31	36	48	$\chi^2(2)= 3.980$	0.137
Mug First	52	12	17	23	$\chi^2(2)= 3.500$	0.174
Chocolate First	63	19	19	25	$\chi^2(2)= 1.140$	0.565

This table reports the mean valuations of the (uniquely) most highly valued goods, by subset. The table pools all treatments *A*-*D* and also separates preferences by first subset of goods.

Mugs	**Squares**	**Circles**	**Triangles**	**Kruskal-Wallis Test**	***p*-value**
All	£3.28	£2.54	£2.81	$\chi^2(2)= 2.701$	0.259
Mug First	£3.51	£2.48	£3.02	$\chi^2(2)= 3.463$	0.177
Chocolate First	£2.85	£2.66	£2.50	$\chi^2(2)= 0.119$	0.942
Chocolates	**RR**	**MC**	**MM**	**Kruskal-Wallis Test**	***p*-value**
All	£2.99	£3.00	£3.02	$\chi^2(2)= 0.006$	0.997
Mug First	£3.02	£2.79	£3.20	$\chi^2(2)= 1.045$	0.593
Chocolate First	£2.97	£3.18	£2.86	$\chi^2(2)= 0.481$	0.786