**Experimental Design**

*Overview*

The experiment had separate buying and selling treatments, faced by different subjects. The buying treatment elicited WTP valuations for a range of consumption goods and lotteries (the *trading commodities*); the selling treatment elicited WTA valuations for the same commodities.  In each treatment, each subject faced eleven tasks in random order, presented on a computer screen. Ten of these tasks had the two-part structure of the canonical design. The first part of such a task was a question that was framed to provide a potential anchor value. Different tasks used different types of anchor, differentiated in terms of plausibility, relevance and engagement. The eleventh task, used as a control, differed from the others in that its first part was a ‘filler’ question with no anchoring significance. The second part of each task elicited the relevant valuation.

The ten non-control tasks faced by any given subject can be grouped into five pairs. (This pairing was not described explicitly to subjects; because of randomization, the two tasks in a pair were usually not adjacent to one another.) In any given pair, the two tasks were identical except that one task provided a relatively low anchor value while the other provided a relatively high one. Thus, our design allows within-subject tests of the existence and size of anchoring effects, in both buying and selling, for each anchor type.

Because we wanted to investigate more than five anchor types but did not want to overload subjects or make the experiment last too long, the subjects in each treatment (i.e. buying and selling) were randomly divided into two groups, A and B. These groups faced different (but overlapping) sets of tasks, involving nine anchor types in all (in addition to the control task). To minimize learning and across-tasks effects, and to ensure that our results were not dependent on the use of any specific commodity, we used six different trading commodities, each with a market value of approximately £5. Each subject’s eleven tasks involved all six commodities (one for each pair of non-control tasks and one for the control task.) To ensure that effects due to differences between anchor types were not confounded with effects due to differences between commodities, anchor types and commodities were counterbalanced.

Because each subject faced eleven tasks rather than just one, we were able to collect a rich body of data and to use within-subject tests. The downside of this design strategy is that the anchor used in one task might influence the valuations reported by subjects in later tasks. If such contamination were to occur, it would add noise to the data. However, because the order of tasks was randomized, it would not impart systematic biases to our tests.

*Anchor types*

The first part of each non-control task was a *comparative question* relating to an *anchor commodity*. Depending on the anchor type, this might or might not be the same as the trading commodity. Using *A* to denote the anchor commodity and £*x* to denote an amount of money, the comparative question took the form ‘If you had *A*, would you sell your *A* if we offered you £*x*?’ (in the selling mode) or ‘If you had £12, would you buy *A* if you had to pay £*x*?’ (in the buying mode). Thus, the subject was prompted to focus on the *anchor value* *x*. In the control task, the comparative question ‘Do you like dogs more than cats?’ was used as a filler. We will say that the control task had the *no lab anchor* type. (We use this term to signal that the ‘anchor value’ provided by the experiment is not the only value that a subject might retrieve from memory when reporting her valuation of the trading commodity.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anchor type | Process to set  anchor value | Relation of anchor commodity to trading commodity | Low anchor value | High anchor value | Incentives | Subject groups facing task |
| Baseline | None | Same | £1–£2 | £10–£12 | No | A and B |
| Implausible price | None | Same | £0.01 | £1000 | No | A |
| Similar good | None | Similar | £1–£2 | £10–£12 | No | A |
| Dissimilar good | None | Dissimilar | £1–£2 | £10–£12 | No | B |
| Incentivized | None | Same | £1–£2 | £10–£12 | Yes | B |
| Passive number search | Computer finds number | Same | £1–£2 | £10–£12 | No | A |
| Passive price search | Computer finds price | Same | £1–£2 | £10–£12 | No | B |
| Active number search | Subject finds number | Same | £1–£2 | £10–£12 | No | B |
| Active price search | Subject finds price | Same | £1–£2 | £10–£12 | No | A |
| No lab anchor | N/A | N/A | N/A | N/A | N/A | A and B |

The anchor types used in the experiment, and the subject groups to which they were assigned, are described in Table 1. The first row of this table describes the *baseline* anchor type, which is similar to that of the canonical design. When this anchor type was used, the subject’s first exposure to the anchor value was when it appeared in the comparative question (indicated by ‘none’ in the ‘process to set anchor value’ column). The anchor commodity was the same as the trading commodity. The low anchor value was drawn at random from the interval from £1 to £2; the high anchor value was drawn at random from the interval from £10 to £12. High and low anchor values were both intended to be perceived as plausible prices or valuations. The baseline comparative question was not incentivized. This was for reasons of external validity*.*  Outside the lab, an anchoring manipulation is typically a way of framing a given decision problem (as when a supermarket prices a product at £6.95, with the label ‘Special offer! Normal price £9.95’); the frame does not have an incentive structure independent of that problem. The entry in the final column indicates that the baseline anchor type was faced by both subject groups. By making two anchor types common to both groups, we were able to check that the randomization was effective and that the particular assignment of anchor types to groups was not inducing systematic effects. In fact, there was no significant difference in WTA or WTP valuations for the common tasks between the two groups.

The other anchor types differed from the baseline in the following ways.

The *implausible price* anchor type was used to investigate the effect of variation along the dimension of plausibility. In this anchor type, the low anchor value was £0.01 and the high anchor value was £1000. We assumed that such extreme values would not be perceived as providing information about (or reference points for) responses to the trading questions in part 2.

The *similar good* and *dissimilar good* anchor types were used to investigate the effect of variation along the dimension of relevance. In these anchor types, the anchor commodity was not the same as the trading commodity, but was approximately equal in market value. In the *similar good* case, the two goods were chosen so that individuals’ ‘true’ valuations of the commodities were likely to be positively correlated. In the *dissimilar good* case, the two commodities were unrelated to one another (see Section 4.4 below).

The *incentivized* anchor type was used to investigate the effect of one form of engagement. In this anchor type, the comparative question was incentivized in the same way as the trading questions (see Section 4.5 below).

The final four non-control anchor types were used to investigate a different form of engagement – involvement in the determination of the anchor value. In each of these anchor types, the comparative task was preceded by a *matrix problem*, whose solution determined the anchor value. In the two *price search* anchor types, the subject was shown an 8×8 matrix of monetary values, described as ‘prices’. These values were determined randomly, subject to the constraint that the lowest value was in the range from £1 to £2 (for low anchor tasks) or from £10 to £12 (for high anchor tasks). In the two *number search* anchor types, the only difference was that the entries in the matrix were dimensionless numbers. (Thus, for example, the ‘number’ 1.45 was shown instead of the ‘price’ £1.45.) In the two *active* anchor types, the subject was asked to find the lowest price (or number) in the matrix and to type it into a blank space on the screen. (If the response was incorrect, the subject was prompted to try again.) This (or the corresponding) price then became the anchor value for the comparative question.

The final row of Table 1 describes the *no lab anchor* control, which was faced by both subject groups.

*The elicitation of WTA and WTP*

The second part of each task began with a screen telling the subject ‘You are endowed with *T* and you have an opportunity to sell *T*’ (in the selling mode) or ‘You are endowed with £12 and you have an opportunity to buy *T*’ (in the buying context). Here *T* denotes the name of the trading commodity. The subject was then asked to answer ‘yes’ or ‘no’ to each of 25 *trading questions* of the form ‘If I am offered £*y* for *T*, I will sell’ (in the selling mode) or ‘If the price of *T* is £*y*, I will buy’. The trading questions used 25 different prices: *y* = 0.01, 0.50, 1.00, 1.50, .... 12.00. Thus, a subject’s responses to these questions located her WTA or WTP within a £0.50 band (or revealed that valuation to be less than £0.01 or greater than £12.00).[[1]](#footnote-1)

Notice that this design elicits valuations by multiple binary choices rather than by a single open-ended question. We used this elicitation method for three reasons. First, ‘Would you pay £x for *T*?’ is cognitively simpler than ‘What is the highest price you would pay for *T*?’ and so less likely to induce confusion. Second, the multiple binary choice format can be linked to the BDM mechanism by telling subjects that one binary choice will be selected at random to be ‘real’. This presentation makes the incentive-compatibility of the mechanism more transparent than when valuations are open-ended. Third, most retail transactions take place at take-it-or-leave-it prices; cases (such as sealed-bid auctions) in which consumers record open-ended valuations or bids are much rarer. Thus, in the context of retail markets, the multiple binary choice format has greater external validity.

As we noted in Section 2, there is some evidence that anchoring effects are weaker when valuations are elicited by binary choices. One possible explanation is that the greater transparency of this method allows subjects to be more confident in their responses and so less susceptible to irrelevant cues. An alternative explanation is compatible with the hypothesis of backward priming. Binary choice questions require yes/no answers while open-ended valuation questions require numerical answers. Thus, numbers in memory are more likely to be accessed when subjects are dealing with open-ended questions.

*Trading and anchor commodities*

Six different trading commodities were used in the experiment: a lottery which all the prizes were positive, which we called a ‘win-win gamble’; five National Lottery scratch cards; two bottles of Chinese sauce; a box of chocolates; a bath towel; and a luxury pen. The win-win gamble gave the prizes £15.53, £3.08 and £0.01 with probabilities 0.3, 0.5 and 0.2 respectively; its expected value was £6.20. The other commodities had market prices in the range from £4.50 to £5.10. (Multiple items, such as five scratch cards, are treated as a single commodity.) In the ‘similar good’ tasks, the corresponding anchor commodities were respectively: a different win-win gamble with approximately the same expected value; five National Lottery scratch cards of a different type; three bottles of Thai sauce; a box of a different type of chocolates produced by the same firm; five face cloths; and a different type of pen. In the ‘dissimilar good’ tasks, the anchor commodity was an iTune coupon.

*Incentives*

At the end of the experiment the computer picked one of the eleven tasks at random. If the anchor type of that task was not ‘incentivized’, the computer then picked one of the 25 trading questions for that task. What the subject took away from the experiment was determined by her response to that task. In a selling task, if the subject had declared her willingness to sell the trading commodity at the £*x* price of the relevant trading question, she received £*x*; otherwise, she received the commodity. In a buying task, if the subject had declared her willingness to buy at the £*x* price, she received the commodity and £(12 – *x*); otherwise, she received £12. If the subject received a win-win gamble, it was resolved by the computer, using a random-number generator. If the anchor type of the task picked was ‘incentivized’, the computer then randomly picked either the first or second part of that task. Depending on which part was picked, the subject’s earnings were determined either by her response to the comparative question or by her response to one of the 25 trading questions.

1. The software was designed so that the subject did not need to click ‘yes’ or ‘no’ to every question. For example, in the selling mode, if a subject clicked ‘yes’ (respectively ‘no’) to an offer of *x*, ‘yes’ (‘no’) was automatically entered for every offer greater than (less than) *x*. Thus only two clicks were needed to answer all 25 questions. This procedure prevented subjects from making inconsistent responses. [↑](#footnote-ref-1)