CeDEx

Discussion Paper

Centre for Decision Research and Experimental Economics

Discussion Paper Series

ISSN 1749-3293

CeDEx Discussion Paper No. 2007–10

Preference Reversals and Disparities between Willingness to Pay and Willingness to Accept in Repeated Markets

Graham Loomes, Chris Starmer and Robert Sugden

November 2007

The University of Nottingham

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of Public Economics, Individual Choice under Risk and Uncertainty, Strategic Interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/economics/cedex/ for more information about the Centre or contact

Karina Whitehead
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15620
Fax: +44 (0) 115 95 14159
karina.whitehead@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/economics/cedex/papers/index.html

Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets

Graham Loomes[*], Chris Starmer[#] and Robert Sugden[*]

*School of Economics, University of East Anglia, Norwich NR4 7TJ, UK

#School of Economics, University of Nottingham, Nottingham NG7 2RD, UK

29 October 2007

Abstract

Previous studies suggest that two otherwise robust ‘anomalies’ – preference reversals and disparities between buying and selling valuations – are eroded when respondents participate in repeated markets. We report an experiment which investigates whether this is true when factors neglected in previous studies are controlled, and which distinguishes between anomalies revealed in the behaviour of individual market participants and anomalies revealed in market prices. Our results confirm the decay of buy/sell disparities, but not of preference reversal. This raises doubts about the hypothesis that, in general, repeated markets reveal anomaly-free preferences, even among the marginal traders who determine prices.

Key words: preference reversal, willingness to accept, willingness to pay, repeated market.

JEL classification: C91 (design of experiments: laboratory, individual behaviour)

Corresponding author: Robert Sugden (e-mail: r.sugden@uea.ac.uk; telephone: +44 1603 593423; fax +44 1603 456259)

There is now a large body of evidence from *individual* decision experiments showing 'anomalies' – that is, departures from standard precepts of rationality – which appear to be substantial, systematic and easily replicable (see e.g. Camerer, 1995; Starmer, 2000). By contrast, many *market* experiments exhibit patterns of behaviour which conform with standard theory. This has prompted interest among experimental researchers in the possibility that anomalies might become less frequent, or disappear altogether, if decision processes were embedded in repeated market environments. A number of such experiments have now been carried out, and in many cases have found a *decay effect*: the anomaly under investigation becomes less frequent and/or or less systematic as individuals gain experience of the market. These results have sometimes been interpreted as supporting the general hypothesis that the anomalies studied by behavioural economists are unimportant in repeated markets (e.g. Plott, 1996). However, it would be premature to accept that hypothesis on the basis of existing evidence.

One reason for caution is that part of the evidence of decay effects comes from experiments featuring arbitrage operations (or 'money pumps') to impose financial losses on individuals who exhibit *preference reversal* (that is, their stated valuations of two alternatives imply a ranking different to that revealed in their direct choice between the same alternatives). We will argue that such transparent, individual-level arbitrage is not a normal feature of most individuals' everyday experience of markets. Hence, to discover whether *actual* market experience reduces the prevalence of anomalies in *actual* market behaviour, it is necessary to investigate the effects of experience in markets in which individual-level arbitrage does not occur.

Although there is a considerable evidence of decay effects in non-arbitraged markets, most of it relates to one particular anomaly: the disparity between willingness to accept (WTA) and willingness to pay (WTP). There is evidence suggesting that some other anomalies may be more robust to market experience. For example, Ariely, Loewenstein and Prelec (2003) find that the tendency for stated valuations of a good to be influenced by arbitrary cues (or 'anchors') persists when valuations are elicited in repeated markets. List (2002) finds a certain kind of dominance violation (the 'more is less' phenomenon) among professional dealers in sportscard markets. If progress is to be made in understanding the role of market experience, it is important to know how general the decay effect is.

An additional reason for caution is that many of the experiments that have found a decay effect in non-arbitraged markets have not controlled for a potentially confounding factor: *shaping effects*. Shaping effects occur when participants in repeated markets revise their subjective valuations of goods in the direction of previously observed market prices, even though those prices have no relevant information content. To the extent that such effects occur, economic agents, contrary to standard assumptions, do not act on preferences that are exogenous to the market process. Much of the evidence for decay effects comes from experiments which have used repeated second-price auction mechanisms. If shaping effects are at work in these mechanisms, mean and median bids in buying auctions will tend to rise with repetition, while mean and median asks in selling auctions will tend to fall. Because of this tendency, apparent decay effects may be artefacts of shaping.

In this paper, we report an experiment which compares the effect of market experience on the WTA/WTP disparity with its effect on preference reversal. Our design allows us to distinguish between the decay of anomalies, as revealed *in the behaviour of individual market agents*, and their decay as revealed *in market prices*. We argue that the most credible theoretical hypothesis about the 'disciplining' effects of market experience has stronger implications for the latter kind of decay effect. By using data on market prices and on the behaviour of the 'marginal' traders whose actions determine prices, our design allows sharp tests of the market discipline hypothesis while controlling for shaping effects.

1. The anomalies, decay effects, and shaping effects

Since Knetsch and Sinden (1984) first reported 'an unexpected disparity in measures of value', a succession of experiments and surveys have found WTA values for goods that are not just marginally higher than the corresponding WTP values – as standard theory would lead us to expect – but exceed them to an extent that cannot credibly be accounted for by that theory.[1] There is a similarly large body of evidence of the preference reversal phenomenon, first reported by Slovic and Lichtenstein (1968). In the classic preference reversal experiment, respondents are presented with two bets – a *$ bet* offering a small chance of a relatively large prize, and a *P bet* offering a larger chance of a smaller prize. Respondents make straight choices between the two bets, and give a certainty equivalent valuation for each of them. Preference reversal is an asymmetric inconsistency between choices and valuations: many respondents *choose* the P bet but *value* the $ bet more highly, while the opposite inconsistency – choosing the $ bet but giving a higher value to the P bet – is relatively rarely observed (for a review, see Seidl, 2002).

The evidence for each of these anomalies has come predominantly from individual choice and/or stated value tasks, but there have been some investigations of the effect of market experience.

In the case of preference reversal, the main focus has been on the effect of exposing individuals to money pumps. Chu and Chu (1990) use a design in which experimental participants whose choices and valuations are inconsistent are compelled to make cycles of trades in which they lose money. Unsurprisingly, when individuals repeat the same choice and valuation tasks and are money-pumped each time they exhibit preference reversal, they quickly learn to report choices and valuations that cannot be exploited. A more interesting question is whether individuals can transfer this learning to new tasks. Chu and Chu find that individuals who have been money-pumped in one set of preference reversal tasks learn to avoid exposing themselves to arbitrage in new preference reversal tasks. Cherry, Crocker and Shogren (2003) find that this learning effect extends to new preference reversal tasks that are merely hypothetical (and so not subject to arbitrage), and in which the outcomes of the lotteries are environmental experiences (such as seeing a grizzly bear); the main effect of learning is to reduce valuations of the $ bet. Cherry and Shogren (2007) find that these effects occur in a 'cheap talk' variant of the design in which subjects are taught about money pumps by hypothetical examples, but not exposed to real losses. They also find that (real or hypothetical) money-pumping of preference reversal tends to reduce individuals' subsequent valuations of low-probability lotteries that are faced outside the preference reversal framework.

These findings are potentially significant for the design of survey instruments for stated preference studies, but they seem less relevant for the explanation of everyday behaviour in real markets. Intuitively, it is easy to understand the sobering effect of an unambiguous loss of money which occurs within a few minutes of the action that gives rise to it. In ordinary experience, however, transparent money pumping is usually found only in confidence tricks, and people's immunity to these probably depends more on simple heuristics (such as not responding to unsolicited doorstep or e-mail offers) than on the cultivation of consistent preferences. Of course, professional arbitrage plays a major role in the determination of market prices, but ordinary consumers experience the effects of such operations only indirectly.[2]

Our main concern is with the effects of repeated trading in *non-arbitraged* markets. Little work has been done to investigate the impact of this form of market experience on preference reversal. One relevant study, by Cox and Grether (1996), finds a reduction in the asymmetry and (to a lesser extent) frequency of preference reversals over the course of an experiment in which valuations were elicited in repeated second-price selling auctions.

In the case of the WTA/WTP disparity, in contrast, there is now considerable evidence that the disparity decays when valuations are elicited using particular market mechanisms and as experience of those mechanisms increases. In several experiments in which valuations have been elicited in second-price Vickrey auctions, the disparity has decayed as the auction mechanism has been repeated (e.g. Coursey, Hovis and Schulze, 1987; Shogren et al, 1994; List and Shogren, 1999). Similar decay effects have been found in repeated median price auctions (Loomes, Starmer and Sugden, 2003) and in repeated random *k*th-price auctions (Shogren et al, 2001).[3] In a field experiment in a market for sportscards, it has been found that the WTA/WTP disparity is least for those participants who have most prior experience of trading in that market (List, 2003).

However, some of this evidence of the decay of anomalies in repeated markets may be due to shaping. Shaping effects take the form of a tendency for participants in repeated markets to revise their bids or asks in the direction of previously observed market prices. Loomes et al (2003) suggest that such effects can occur if, before an individual confronts an elicitation mechanism, his or her preferences are imprecise or not fully articulated. In such a setting, responses may be generated using heuristics in which market prices act as cues. This can occur even if values are entirely private and so prices convey no information which should be expected to influence those values. In a second-price auction with more than three participants, the market price is a biased indicator of median bids or asks: it is higher than the median bid in a buying auction and lower than the median ask in a selling auction. Thus, shaping effects induce increases in median bids and decreases in median asks. Clearly, if the WTA/WTP disparity is measured by differences between median asks and bids in second-price auctions, the decay of this disparity may be a shaping effect.

A similar mechanism can also generate apparent decay of preference reversal in designs in which certainty equivalents are elicited in second-price selling auctions. Because the probability of winning a P bet is high, there is a narrower interval of 'reasonable' or 'credible' values than is the case for $ bets, where the considerably larger prize places a higher upper bound on valuations and allows a greater dispersion of responses. Thus, if preferences are imprecise, shaping effects are likely to have a greater impact on valuations of $ bets than on those of P bets. Since, in second-price selling auctions, these effects tend to shift the distribution of valuations downwards, the strongest effect will be a reduction in the valuations of $ bets, which will tend to reduce the incidence of preference reversal.

Do shaping effects occur in fact? List and Shogren (1999) analyse data from a set of second-price auctions for sandwiches with varying degrees of risk of containing pathogens. They find that shaping effects are significant but small in magnitude, and are absent in treatments in which participants are given fuller information about the goods being traded. Loomes et al (2003) report direct evidence of shaping effects in repeated Vickrey auctions for lotteries. Knetch, Tang and Thaler (2001) report what appears to be a related effect. In repeated ten-person second-price auctions for coffee mugs, WTP increases and WTA falls; but in repeated *ninth* price auctions – in which prices understate median bids and overstate median asks – WTP *falls* and WTA *increases.* In the present context, it is particularly significant that shaping effects were present in Cox and Grether's experiment on preference reversal: in the second-price auctions, asks were significantly and positively related to previously-observed market prices (Cox and Grether, 1996, p. 400).

2. Market discipline

If we are to test the hypothesis that market experience reduces the incidence of anomalies, we first need to formulate that hypothesis more precisely. As usually stated, the hypothesis proposes that individuals have anomaly-free *underlying preferences*, which are independent of the institution in which preferences are elicited. Facing unfamiliar tasks, individuals may be susceptible to various errors and biases; but, as a result of exposure to the incentives and feedback generated by repeated markets, such errors and biases are increasingly filtered out.

But what exactly is the filtering mechanism? First, it is important to distinguish between *repeated markets* and *repeated consumption*. Our concern in this paper, like that of most of the literature we have cited, is with the former.[4] That is, we are concerned with the effects of feedback within a market institution. In the market institutions investigated by experimental economists, the feedback received by an agent typically takes the form of information about the market price and about whether or not his bid or ask has been accepted. Following Loomes et al (2003), we suggest that the most credible filtering mechanism within such institutions is *market discipline* – the punishment of error. The market discipline hypothesis proposes that, as a market is repeated, individuals tend to correct those errors that, in previous markets, have proved costly.

To see the implications of this hypothesis, consider a repeated *k*th price Vickrey buying auction with *n* participants. (The case of a *k*th price selling auction can be analysed in a symmetrical way.) Consider any round of this auction, other than the last. Each participant *i* submits a bid b_i. For simplicity, suppose that no two bids are the same, and let the participants be indexed so that $b_1 > b_2 > ... > b_n$. The market price is set at b_k, and persons 1, ..., $k - 1$ buy at this price. For each person *i*, an *underlying valuation* v_i can be defined as the valuation that reflects her underlying preferences; if $b_i > v_i$, we will say that she *overbids*, while if $b_i < v_i$ she *underbids*. There are exactly two forms of costly error that *i* can make. The first is that she buys when the price is higher than her underlying valuation; this is the case if and only if $b_i > b_k > v_i$. In this case, the market discipline hypothesis implies that, having experienced a loss as a result of overbidding, she will reduce her bid in the next round of the auction so that $b_i > b_i' \geq v_i$, where b_i' is *i*'s bid in the next round. The second form of costly error is that she fails to buy when the price is lower than her underlying valuation; this is the case if and only if $v_i > b_k \geq b_i$. Having missed out on an opportunity for gain through underbidding, she will increase her bid in the next round so that $v_i \geq b_i' > b_i$.

Notice that there are other ways in which person *i*'s bid may differ from her underlying valuation, but which are *not* punished. For example, if $b_k > b_i > v_i$, she overbids but does not buy. If $b_i > v_i > b_k$, she overbids and buys, but the price is less than her underlying valuation. If $v_i > b_i > b_k$, she underbids but still buys. If $b_k > v_i > b_i$, she underbids and fails to buy, but the price is greater than her underlying valuation. In such cases, the market discipline hypothesis does not predict any systematic change in the individual's bid.

We will say that the vector (b_1, ..., b_n) of bids is in *equilibrium* if no person's bid leads to a costly error. Although we are not using an explicitly game-theoretic model, this equilibrium concept is essentially that of Nash equilibrium: a vector of bids is in equilibrium if each individual bid is optimal for the bidder, given the bids of the other participants. Clearly, the market discipline hypothesis implies that, as long as bids are not in equilibrium, *some* bids will change, and each such change will take the relevant person's bid closer to her underlying valuation. Recall that, by assumption, underlying valuations reflect anomaly-free preferences. Thus, aggregating across the behaviour of all market participants, anomalies will become less pronounced as markets are repeated, until equilibrium is reached. Since the hypothesised mechanism is one of error-correction, we should expect deviations from the standard rationality assumptions to become less frequent – and not merely less asymmetrical – as markets are repeated.

However, the market discipline hypothesis does *not* imply that, after sufficient repetition, all market participants reveal their underlying preferences in their bids. To the contrary, in equilibrium there can be either or both of overbidding and underbidding. Equilibrium has two defining features. First, for each individual $i = 1, \ldots k - 1$, $v_i \geq b_k$. (That is, no individual who buys at the price b_k has reason to regret doing so.) Second, for each $i = k, \ldots, n$, $b_k \leq v_i$. (No individual who fails to buy at this price has reason to regret not buying.) Thus, each individual's decision about whether or not to buy at the market price reveals either a lower bound (if he buys) or an upper bound (if he doesn't buy) to his underlying valuation. It follows that, in equilibrium, the market price is no greater than the $(k - 1)$th highest underlying valuation, and no less than the kth highest underlying valuation. Thus, in a kth price buying auction, the equilibrium price is approximately equal to the kth highest underlying valuation. A symmetrical analysis shows that, in a kth-price selling auction, the equilibrium price is approximately equal to the kth lowest underlying valuation. In other words, the market discipline hypothesis implies that, after sufficient repetition, *the market price reveals the underlying valuation of the marginal trader*. Market discipline filters out anomalies in so far as they impact on market prices and on the trading decisions that are made at those prices.

In the light of this analysis, we can distinguish between two strategies for testing the market discipline hypothesis. The *individual-based* strategy is to collect data on the bids or asks of all participants in a sequence of repeated markets, and to investigate whether the frequency and asymmetry of anomalies, as revealed in those bids or asks, decays with repetition. The market discipline hypothesis predicts such a decay, but it does not predict that, in the limit as the market is repeated many times, anomalies will be eliminated altogether. The *market-based* strategy is to collect data on market prices (or, equivalently, on the bids or asks of marginal traders), and to investigate whether the frequency and asymmetry of anomalies, as revealed in those prices, decays with repetition. In this case, the market discipline hypothesis predicts not only that there will be such a decay but also that, in the limit, anomalies will be eliminated.

In Section 4, we will discuss how these two strategies can be used in relation to WTP/WTA disparities and preference reversal. First, however, we consider how to control for shaping effects in tests of the market discipline hypothesis.

3. Controlling for shaping effects

We now propose two complementary methods of controlling for shaping effects. The first responds to a particular mechanism, discussed in Section 1, by which shaping effects can confound tests of the market discipline hypothesis. The market discipline hypothesis implies that bids and asks tend to move towards underlying valuations, while the shaping hypothesis implies that they tend to move towards previously observed market prices. If the anomaly under investigation is a disparity between bids in buying markets and asks in selling markets, second-price auctions (with more than three participants) generate systematically asymmetric shaping effects: the market price in a buying auction is taken from the upper end of the distribution of bids, while that in a selling auction is taken from the lower end of the distribution of asks. An obvious way to remove this source of bias is to use median price auctions. This ensures that, in both buying and selling markets, the market price is the median of the distribution of bids or asks. Our experimental design uses median price auctions.

However, there is still a problem of distinguishing between changes in bids or asks that are due to market discipline and changes that are due to shaping. For example, suppose we observe that, with repetition, the central tendency of the distribution of bids or asks stays constant, but the variance of the distribution falls. This is consistent with the market discipline hypothesis, on the assumption that the initial distribution of bids or asks reflects random errors. But, given the median price rule, it is also consistent with the shaping hypothesis.

The simplest way to control for shaping is to investigate movements in market prices. This method exploits an important characteristic of shaping effects: these effects do not impinge on market prices. To see why not, reconsider the case of a repeated kth-price buying auction with n participants. In some given round, each participant i submits a bid b_i, participants being indexed so that $b_1 > b_2 > ... > b_n$. The market price is set at b_k. According to the shaping hypothesis, bids in the next round are adjusted in the direction of the market price. Let b_i' be i's bid in the next round. For $i = 1, ..., k - 1$, we have $b_i > b_i' > b_k$; for $i = k$ we have $b_i' = b_i$; and for $i = k + 1, ..., N$, we have $b_i < b_i' < b_k$. Thus, the kth highest bid in the next round is b_k'; since $b_k' = b_k$, shaping has had no effect on the price. Equivalently, it has had no effect on the bid of the marginal trader k. All that shaping has done is to induce *convergence* of non-marginal bids towards the marginal bid. A symmetric analysis applies to selling auctions. The implication is that, by using market-based tests of the market discipline hypothesis, we can control for shaping effects.

4. Experimental design: principles

Our experiment elicited WTA and WTP valuations of, and choices between, two lotteries, one of which had the characteristics of a $ bet, the other those of a P bet. Valuations were elicited in repeated median-price Vickrey auctions. Choices were made in the final stage of the experiment, after participants had taken part in all the auctions.

This design allows us to explore the effect of market experience on WTA/WTP disparities for two different lotteries, and on two kinds of preference reversal: *buying reversals* (where valuations are expressed as WTP) and *selling reversals* (where they are expressed as WTA). For each of these anomalies, we investigate the effects of market experience both at the level of the individual and at the level of the market. This requires us to be able to identify anomalies in the data generated by a Vickrey auction. At the individual level, this is straightforward. We now consider how anomalies can be identified at the market level.

First, consider WTA/WTP disparities for a given lottery. The benchmark for measuring anomalies is given by the null hypothesis that each participant has an underlying valuation for the lottery, and (on the assumption that income effects are negligible) that this valuation is the same, whether it is expressed as WTA or as WTP. There is an anomaly in market prices to the extent that they diverge in some systematic way from what they would have been, had each individual recorded an anomaly-free underlying valuation. If the null hypothesis is true, for any given group of individuals, the median valuation of the lottery is well-defined. If these individuals are the participants in a median price buying auction, and if each individual's bid is equal to his valuation, the market price will equal the median valuation. Similarly, if they are the participants in a median price selling auction, and if each individual's ask is equal to his valuation, the market price will equal the median valuation. Thus, systematic differences between the market prices generated in the two types of auction can be treated as anomalies.

Now consider preference reversal with valuations elicited in median price selling auctions. (A symmetrical analysis applies to buying reversals.) Our null hypothesis is that each individual has anomaly-free underlying preferences, that income effects are negligible, that each individual's asks are equal to her underlying valuations of the relevant lottery, and that each individual's choice reveals her underlying preference between the two bets. We interpret 'anomaly-free' to imply 'consistency with the axioms of expected utility theory'. For any group containing an odd number of individuals, we can use median price selling auctions to generate market prices for $ and P. By confronting each individual with the task of choosing either $ or P, we can discover which bet is chosen by the majority of individuals. It is natural to say that a *classic* preference reversal occurs at the market level if the market price for $ is higher than that for P, but P is chosen by the majority of participants, and that a *counter* reversal occurs if P has the higher market price but $ is the majority choice.

If we are to treat market-level preference reversals as anomalies, we need to show that they are inconsistent with our null hypothesis. That hypothesis allows us to say, of each individual, whether he has an underlying preference for $ or P. (For simplicity, we assume that no one is exactly indifferent.) We need to show that, given the null hypothesis, if the majority of individuals prefers $ to P, then $ has the higher market price; and conversely. In the Appendix, we show that this is an implication of expected utility theory, for any utility function that can be approximated by the first three terms of a Taylor series.

Although our experimental design uses the same data on bids and asks to investigate WTA/WTP disparities and preference reversals, the two types of anomaly are independent in the sense that the presence or absence of one does not imply the presence or absence of the other. The WTA/WTP disparity is a property of comparisons *between bids and asks* for a particular bet; so these comparisons can be made separately for each of the two bets. For given data on choices, preference reversal involves comparisons *between valuations* of the two bets, elicited under the same market type; so these comparisons can be made separately for bids (i.e. testing for buying reversals) and asks (i.e. testing for selling reversals).

However, there may be some overlap between the causal mechanisms that generate the anomalies. Given the hypothesis that people have anomaly-free underlying preferences, the WTA/WTP disparity is possible only if there is underbidding and/or overasking. Because the range of 'reasonable' valuations is greater for $ than P (see Section 1), the extent of both overasking and underbidding can be expected to be greater for $ than for P. Thus, a tendency to overask is likely to increase the prevalence of classic *selling* reversals (by raising WTA valuations of $ more than WTA valuations of P).[5] In contrast, a tendency to underbid is likely to *reduce* the prevalence of classic *buying* reversals (by lowering WTP valuations of $ more than WTP valuations of P).[6] One implication of this is that, if the WTA/WTP disparity decays with market experience, some decay of classic selling reversals – and some countervailing effect for buying reversals – may occur as a by-product. However, by studying both selling and buying reversals, we can investigate whether there is a decay effect for preference reversal *in general*.

In our experimental design, each respondent makes only one choice between P and $. At first sight, it might seem that this feature of our design treats choices and valuations asymmetrically: the valuation tasks are repeated, while the choice task is not. But recall that our objective is to investigate the effects of experience *of market institutions*. Each valuation task is implemented in a Vickrey auction. A Vickrey auction is a market institution in which each participant reports a bid or ask, and receives feedback in the form of information about the market price and about whether she has bought or sold the relevant lottery. We can investigate a well-defined hypothesis – the market discipline hypothesis – about how feedback from earlier rounds of an auction affects behaviour in later rounds. In contrast, a pairwise choice between two lotteries is not a market institution in this sense. Given that there is no consumption during the experiment, a pairwise choice task can provide no new information as feedback: a participant simply records which of two lotteries she prefers. However, by placing the choice task at the end of the experiment, we allow maximum scope for experience gained in trading the P and $ bets to impact on choices. In fact, as we will explain later, we do not need to assume that, were the choice task repeated, responses would remain constant.

5. Experimental Design: Implementation

A total of 175 individuals took part in the experiment, drawn from the general student population at the University of East Anglia. At the start of each experimental session, each respondent was randomly assigned to a *trading group* of either five or seven people; there were 33 trading groups in all.[7] Each session began with an explanation of the auction procedures, structured around practice auctions involving induced value vouchers.[8] After taking part in two such practices, each trading group then took part in two real voucher auctions – one buying, one selling. Next, the special features of lottery auctions were explained, and there were two practices of such auctions, followed by eight lottery auctions. The four auctions which are the focus of this paper – a buying and a selling auction for each of $ and P – were mixed in with four other lottery auctions[9] and were presented in random order. This was intended to minimise any systematic tendency for the last round of a buying (selling) auction for a particular bet to shape the first round of the selling (buying) auction for the same bet. In buying auctions, each respondent was endowed with £9 in cash; in selling auctions, each respondent was endowed with the relevant bet. Once all auctions had been completed, respondents were presented with a pairwise choice between $ and P. The $ bet offered a 0.19 probability of winning £18 (and a 0.81 probability of winning nothing); the P bet offered a 0.81 probability of winning £4 (and a 0.19 probability of winning nothing).[10] Thus, the expected value of the $ bet (£3.42) was slightly higher than that of the P bet (£3.24).[11]

Each auction was repeated six times in immediate succession. In each round of every auction, bids or asks were elicited through an interactive computer program. We designed the elicitation procedure to be as simple and transparent as possible; it was developed and refined through pilot experiments. A guiding principle for the design, widely accepted among stated preference researchers, is that individuals cope more easily with conditional questions requiring yes/no responses (‘If the price was *x*, would you buy?’) than with unconditional open-ended questions using maximum or minimum concepts (‘What is the highest price you would pay?’). Price-conditional questions are less likely to prompt misplaced strategic reasoning and bargaining heuristics. They also facilitate understanding of how stated bids or asks combine with the price to determine outcomes for the respondent.[12]

The elicitation procedure used a fixed set *X* of possible values of the market price. For both bets, and for both buying and selling auctions, we used the set *X* = {£0.01, £0.50, £1.00, ..., £8.50, £9.00}.[13] For each price in this set, the respondent stated, either directly or indirectly, whether she would trade at that price. Consider a buying auction. The program asked each respondent a series of questions of the form: ‘Would you pay £*x*?’, adjusting the value of *x* according to her previous answers. In the first question, the value of *x* was £9.00 (i.e. the highest possible). If the respondent was willing to pay this amount, the procedure halted. If not, the value of *x* was successively reduced to £6.00, £3.00 and £0.01, until she was willing to pay the price stated; if she refused to pay £0.01, the procedure halted. If, say, she reported willingness to pay £6.00 on the second question, the value of x in the third question was £8.50, and was successively reduced by increments of £0.50 in the following questions, until she reported willingness to pay. At the end of the sequence, the computer summarised the implications of the respondent's answers in the form: 'You have said you are willing to pay x' but you are not willing to pay x''', where x' was the largest amount the respondent had said she would pay and x'' was the smallest amount she had said she would not pay. The computer then asked the respondent to confirm that she was happy with this statement. If she did not confirm, the elicitation procedure recommenced. If the respondent confirmed, the task ended; x'' was then recorded by the program as her 'just not willing to trade' value. If she did not confirm, the elicitation procedure recommenced.

An analogous procedure was used in selling auctions, using the same conditional sequence of values of x. In this case, the first question asked if the respondent would accept £9.00; if so, she was asked if she would accept £6.00, and so on. Because of this symmetry between the elicitation of bids and asks, and because the elicitation procedure is exactly the same for P and $ bets, observed WTA/WTP disparities or preference reversals cannot be explained as artefacts of the procedure.

When analysing responses and reporting results, we set a respondent's recorded valuation of the relevant lottery at half-way between x' and x'': for example, a respondent who stated in a buying auction that she was willing to pay £3.00 but not £3.50 is treated as recording a valuation of £3.25; correspondingly, a respondent who stated in a selling auction that she was willing to accept £3.00 but not £2.50 is treated as recording a valuation of £2.75.

After each round of each auction, the market price for that round was determined. Respondents were told the market price for that round and its trading implications for them (i.e. whether they had bought or sold the bet, and if so, how much they would pay or receive). By virtue of the median price rule, the market price in a buying (selling) auction was the median of the 'just not willing to buy (sell)' values reported by participants in the relevant market. Bets were not played out at this stage. Each respondent knew from the outset that one round of one auction (or else the pairwise choice task) would be selected at random at the end of the experiment, and that whatever decisions had been made in the selected round or task would be implemented. Each experimental session lasted about 60 minutes, and the average payment per respondent was £7.06.

6. Results: WTA/WTP Disparities

Table 1 reports summary statistics concerning market prices in the six rounds of the four auctions; the means are graphed in Figure 1. For ease of comparability, we convert market prices into 'valuations by median traders' by adding £0.25 (for selling auctions) or subtracting £0.25 (for buying auctions).[14] Throughout the discussion of the results, the term 'median trader' refers to the participant (or, in the case of a tie, one of the participants) in each auction who reported the median valuation, and hence set the price; thus, an analysis of the valuations reported by median traders is equivalent to an analysis of market prices. These data suggest that, by the end of the sixth round of each auction, market prices had stabilised. In the remainder of the paper, we investigate the effects of market experience by comparing Rounds 1 and 6. In this Section, we consider WTA/WTP disparities.

First, we look at the data at the individual level. Table 2 reports mean, median and standard deviation for asks and bids for each bet in each of Rounds 1 and 6. For both bets, mean asks and bids fell as the market was repeated, but the average fall in asks was markedly greater than the corresponding fall in bids. Mean asks remain greater than mean bids for both bets in Round 6. Notice that mean bids and asks for the $ bet are above the bet's expected value (£3.42) in both Round 1 and Round 6, while the opposite is true for the P bet (whose expected value is £3.24).

Table 3 reports the distribution of differences between asks and bids for each bet in each of those two rounds. (Differences are defined to be positive when asks are greater than bids.) There was a definite tendency for these differences to reduce over the course of the experiment, but they remained significantly positive for both bets in Round 6 ($p < 0.01$ in one-tailed t-tests). The observation that market experience reduces but does not eliminate anomalies at the individual level is consistent with the market discipline hypothesis.

Table 4 presents the corresponding data at the market level. In Round 1, median traders' asks for each bet are higher than their bids. This is most pronounced for the $ bet, but as the right hand column of the table shows, the mean disparity is significant for both bets ($p < 0.01$ in one-tailed t-tests). Between the two rounds, median traders' asks and bids fall for both bets, but asks fall more than bids (see Table 1). By Round 6, although mean disparities remain positive for both bets, neither is significantly different from zero. This pattern is consistent with the market discipline hypothesis, which predicts that anomalies at the market level are eliminated by experience. Viewed in relation to that hypothesis, however, it is perhaps surprising that the degree of dispersion in the distributions of ask/bid differences – at both individual and market level – changes very little between the two rounds. This observation suggests that there may be some source of stochastic variation in preferences which is not ‘corrected’ by market discipline.

5. Results: Preference Reversals

Our findings in relation to the WTA/WTP disparity are broadly in line with the developing consensus that this anomaly decays with market experience. Is there evidence of a similar decay in the case of preference reversal?

We begin by looking at the individual-level data for all 175 participants. Table 5 reports the number of observations of each of the six possible combinations of individual choice and relative valuation for both selling and buying tasks in Round 1 and in Round 6. The bold numbers in the middle columns are the ones relevant to the preference reversal phenomenon. Using valuations from the first round of the selling market, 49 of the 175 participants gave the classic preference reversal responses (choosing the P bet but valuing the $ bet more highly), while only 10 gave the counter reversal responses (choosing the $ bet but valuing the P bet more highly). If instead we use valuations from the first round of the buying auctions, the corresponding numbers are 39 classic reversals and 10 counter reversals. In the Round 1 data, the asymmetry between the two forms of reversal is highly significant for both buying and selling valuations ($p < 0.01$ in both cases, using a one-tailed binomial test with the null hypothesis that the two reversals are equally probable).[15]

Between Round 1 and Round 6, mean asks and bids for both bets fell, but by far the largest fall was in asks for the $ bet (see Table 2). As a result, the proportion of participants who gave higher selling valuations to the $ bet than to the P bet fell between the two rounds. This led to some reduction in the asymmetry between classic and counter reversals for selling valuations, but in Round 6 that asymmetry remained large (42 classic reversals compared with 19 counter reversals) and highly significant ($p < 0.01$ in a binomial test). There was no reduction (in fact, a slight increase) in the total frequency of reversals. In the case of buying valuations, the frequency and asymmetry of reversals was almost unchanged (41 classic reversals and 12 counter reversals). In short, we found little evidence that market experience reduces the general prevalence of preference reversals, and what evidence we did find was confined to selling reversals.

Table 6 reports the corresponding market-level data. At this level, unfortunately, the number of reversals, even in Round 1, is too small to allow useful statistical tests. This is due partly to the relatively small number of market-level observations, and partly to the high proportion of $ choices in the sample. That said, there is a marked fall in the frequency and asymmetry of selling reversals between Round 1 and Round 6; but there is no evidence of any decay effect for buying reversals.

An alternative way of approaching the market-level data is to look at the trends in the valuations of median traders, shown in Table 1. If market experience is to produce a decay in preference reversal, it can do so only by lowering the market price of $ relative to that of P. Such an effect is visible in selling markets (where the ratio of the marginal-trader valuations of the two bets falls from 1.62 in Round 1 to 1.27 in Round 6), but not in buying markets (where the ratio is 1.27 in both rounds).

Yet another way of trying to isolate the effect of market experience on marginal traders is to use individual-level data for subsets of respondents, defined by the closeness of their bids or asks to the market price. For the purposes of this analysis, we define a bid or ask to be *marginal* in a given market if the valuation it implies is within £0.50 of the price. For example, consider a buying market in which the price is £4.00. An individual who reports that she is willing to pay £4.00 but not £4.50 has made a marginal bid: she is *just* willing to pay the market price. Symmetrically, someone who reports that he is willing to pay £3.50 but not £4.00 has also made a marginal bid: he is *just not* willing to pay the market price. The market discipline hypothesis implies that, after sufficient repetition, the market price reveals the underlying valuation of the marginal trader. Given this hypothesis, one should also expect that, after repetition, marginal bids and asks are more reliable than non-marginal ones as indicators of the underlying preferences of the relevant individuals. Thus, the tendency for market experience to eliminate anomalies should be particularly marked in relation to individuals who make marginal bids and asks.

Table 7 reports the incidence of preference reversals for those individuals who, in a given round, made marginal bids *both* in the market for the $ bet *and* in the market for the P bet. Notice that, for both buying and selling, the number of individuals making marginal bids or asks increases between Round 1 and Round 6. This is consistent both with market discipline (i.e. as the result of a reduction in random error) and with shaping (i.e. as the result of a tendency for bids and asks to converge on the market price). However, the asymmetry between classic and counter reversals is statistically significant in both rounds, both for buying and for selling ($p < 0.05$ in binomial tests). Comparing the Round 6 data in this table with those in Table 5 (which refer to all individuals), there seem to be no systematic differences in the patterns of reversals. (For example, of all 175 individuals, 24 per cent committed classic selling reversals in Round 6, while 11 per cent committed counter reversals. The corresponding figures for the 58 individuals making marginal asks are 29 per cent and 8 per cent.)

So far, our analysis of preference reversal has focused on the effects of market experience on *valuations*. As we explained in Section 4, the concept of 'market experience' is not easily applicable to *choice* tasks, since such tasks do not provide new information as feedback. Our measures of changes in the frequency of preference reversal have been based on changes in the distributions of valuations, assessed relative to participants' responses to a single choice task located at the end of the experiment. In appraising our results, however, it is legitimate to ask whether our conclusions might have been different if the choice task had been repeated sufficiently many times. Might such repetition have led to an array of choices which, when combined with the Round 6 valuations, would have revealed a decay effect for preference reversal?

We think this unlikely, for two reasons. First, several studies have confronted experimental participants with long sequences of pairwise choices between lotteries, some choice tasks being faced more than once. These studies have found that choices tend to become more risk-averse as the experiment progresses (Hey and Orme, 1994; Ballinger and Wilcox, 1997; Loomes, Moffatt and Sugden, 2002).[16] So it is reasonable to conjecture that, had the preference-reversal choice task been repeated, the frequency of P-bet choices would at least have remained constant and possibly increased. If valuations are held constant, switches of choice from $ to P unambiguously *increase* the frequency of classic preference reversals.[17]

Second, recall that valuations of the $ bet – in both selling and buying markets, and both for median traders and for participants in general – were persistently higher than the bet's expected value. In itself, this result is indirect but powerful evidence of an inconsistency between valuation and choice. One of the most robust findings of experimental investigations of pairwise choice between lotteries is that individuals are predominantly risk-averse. Since risk aversion (or, in the limit, risk neutrality) is the default assumption in the received theory of choice under uncertainty, it would be very odd to propose that risk-averse choices are anomalies, and that if individuals accumulated sufficient experience, their choices would become risk-loving. But if choices are risk-averse, the $-bet valuations in our experiment are anomalous. To see the nature of the anomaly, let *V*($) be some individual's WTA valuation of the \$-bet, let E(\$) be the bet's expected value, and assume $V(\$) > E(\$)$. Now consider any sum of money x such that $V(\$) > x > E(\$)$. If the individual's choices are risk-averse, she will prefer the certainty of x to the \$ bet in a pairwise choice, even though her valuation of the bet is greater than x. This phenomenon, sometimes referred to as 'overvaluing' the \$ bet, is a form of preference reversal.[18] Tversky, Slovic and Kahneman (1990) have shown that a large proportion of classic preference reversals are associated with overvaluation of the \$ bet. Recall too that in the experiments of Cherry et al (2003) and Cherry and Shogren (2007), the main effect of applying money pumps to individuals who commit preference reversals is to reduce valuations of the \$ bet (see Section 1 above). It seems that, if any learning mechanism is to eliminate preference reversal, it must impact on the valuations of \$ bets.

Taking all the evidence together, we find little support for the hypothesis that preference reversal is eroded by experience of (non-arbitraged) markets.

6. Discussion

How can we reconcile what appear to be conflicting patterns in the data? The market discipline hypothesis seems to perform well in relation to the WTA/WTP disparity, but rather poorly in relation to preference reversal. Buying reversals in particular seem to be robust to the feedback provided by median price Vickrey auctions. Why might the two anomalies differ in their susceptibility to erosion by market experience? In thinking about this issue, a natural starting point is to consider why the two anomalies occur at all.

In the case of the WTA/WTP disparity, most of the explanations that have been proposed involve one or other of two types of mechanism. One is *strategic bidding (and asking)*: individuals who fail to understand the incentive-compatible properties of Vickrey auctions tend to understate their true valuations when submitting bids and to overstate their true valuations when submitting asks. These under- and over-statements may be deliberate (although mistaken) attempts to gain an advantage, or they may result from the use of heuristics which are adapted to everyday bargaining problems. The second mechanism is *loss aversion*: individuals tend to give greater weight to something they give up than to a gain of the same objective magnitude, and thus are disproportionately averse to moving away from their initial endowments.

As we noted in Section 4, strategic bidding and loss aversion may each contribute to preference reversal when valuations are elicited in selling tasks (but will tend to act against buying reversals). However, most of the hypotheses that have been proposed as explanations for preference reversal apply equally to buying and selling reversals; they refer to differences between the mental processes that are brought into play by valuation tasks and by choice tasks. These mental-processing hypotheses have no implications for WTA/WTP disparities, which are revealed in comparisons between one valuation task and another.

One such hypothesis is that of *scale compatibility*, proposed by Tversky, Sattath and Slovic (1988). The general idea is that the response mode for a task may be more or less compatible with the various attributes of objects under consideration in the task; the more compatible an attribute is with the response mode, the more weight it has in determining the individual's response. In the case of preference reversal, 'probability of winning' and 'amount to win' are attributes of bets; of the two bets, the P bet is better in terms of the first attribute while the $ bet is better in terms of the second. In valuation tasks, the response mode (a money scale) is particularly compatible with the 'amount to win' attribute, and so the $ bet is favoured in valuation tasks.[19]

An alternative (but closely related) explanation is that, in dealing with valuation tasks, respondents use *anchoring* heuristics. The idea here is that when valuing a bet, individuals anchor on the money payoff and then adjust downwards to allow for the fact that the chance of getting this payoff is less than 1. If, as much psychological evidence suggests, the anchor exerts undue weight and the downward adjustment is insufficient, this heuristic will tend to overvalue the $ bet relative to the P bet: the required adjustment, and hence the potential for insufficient adjustment, is much greater for the $ bet.

A third mental-processing explanation is based on the hypothesis of *imprecision* in individuals' preferences and valuations (Butler and Loomes, 2007). An individual who is asked to state a certainty equivalent value of a lottery may be able to identify a range of amounts of money that are perceived to be *possible* equivalents, but may find it hard to pick out a single value from that range. Because there is a high probability of winning the P bet, the range of possible money equivalents is relatively narrow, bounded above by the value of the prize. The corresponding range for the $ bet tends to be wider and to involve higher values, even among individuals who would choose the P bet in a straight choice. If each individual's reported valuation of each bet is picked randomly from the corresponding range, classic preference reversals will be induced with greater frequency than counter reversals.

Since these mental-processing mechanisms are distinct from both strategic bidding and loss aversion, there is no direct inconsistency in the hypothesis that the WTA/WTP disparity is more susceptible to erosion by market experience than preference reversal is. We now offer some tentative suggestions as to why that hypothesis might be true.

The strategic bidding mechanism can work only if respondents misunderstand the market institution in which valuations are being elicited. Thus, there is an obvious reason to expect the propensity for strategic bidding to decrease as respondents learn more about how the market institution works. The market discipline hypothesis offers a credible account of how participants might learn to avoid making strategic bids or asks which lead to costly errors. But the heuristics by which, according to the mental-processing hypotheses, respondents arrive at valuations are not strategic attempts to take advantage of ill-understood market institutions. They are methods by which individuals who do not have pre-existing and well-defined valuations construct valuations in responses to tasks which require them to do so. Such heuristics may continue to be used by respondents who understand the market institution in which they are placed.

At first sight, loss aversion might seem less obviously susceptible to erosion by market experience. Nevertheless, there seems to be at least one route by which market experience might impact on loss aversion. There is considerable experimental evidence that reference points are labile: for example, respondents often exhibit loss aversion about giving up trivial quantities of consumer goods (a coffee mug, a bar of chocolate) that they have received only a few minutes before. Given this fact, it may be that repeated experience of trading a particular kind of good – perhaps even just of forming plans to buy or sell it, conditional on the price falling into some specified range – weakens the individual's perception of 'not trading' as a salient reference point. If this conjecture is correct, repeated participation in a market will tend to erode loss aversion and the WTA/WTP disparity that it generates.

However, even if both strategic bidding and loss aversion are eroded by market experience, the values to which marginal traders' bids and asks converge may still be conditioned by their having been generated in valuation tasks: they may still carry the imprint of whatever heuristics are brought into play by such tasks. Of course, this could not be the case if individuals had well-defined, context-independent underlying preferences which they 'discovered' through participation in markets. But recall our finding that, although market experience reduces *mean* differences between asks and bids, it does not reduce the degree of *dispersion* of ask/bid differences (Section 6). The implication is that individuals' valuations of the two lotteries are imprecise. Market experience does not erode this imprecision; it merely brings the range of 'possible' WTA valuations of each lottery into alignment with the corresponding range of WTP valuations for the same lottery. If individuals do not have context-independent underlying preferences so that their valuations are constructed by the use of contextual heuristics, and if this results in 'overvaluation' of $ bets relative to P bets, there seems to be nothing in the market experience *per se* that will tell them that those relative valuations require revision.[20]

Our main conclusion, then, is that 'anomalies' cannot be treated as a homogeneous category. Our results suggest that some anomalies are eroded by market experience, while others are not. A subsidiary conclusion is the importance of distinguishing between anomalies at the level of market prices and anomalies at the level of bids and asks: even if market discipline eliminates WTA/WTP disparities in market prices, such disparities may persist in the bids and asks of non-marginal traders. However, although we have offered some conjectures about why market experience may erode some anomalies and not others, this question remains open: there is, as yet, no well-developed theory of the mechanisms by which particular characteristics of markets interact with particular characteristics of anomalies. We hope that the development and testing of such theories will be prominent on the research agenda of the next few years.

Table 1: Valuations reported by median traders (in £; *n* = 33)

Sell $

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
mean	5.40	4.66	4.25	4.07	3.87	3.83
median	5.25	4.25	4.25	3.75	3.75	3.75
standard deviation	1.34	1.19	1.21	1.39	1.43	1.38

Buy $

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
mean	3.75	3.75	3.63	3.81	3.89	3.64
median	3.75	3.75	3.25	3.75	3.25	3.75
standard deviation	1.08	1.02	1.01	0.88	0.95	0.91

Sell P

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
mean	3.34	3.22	3.08	3.07	2.98	3.02
median	3.25	3.25	3.25	2.75	2.75	2.75
standard deviation	0.52	0.71	0.62	0.57	0.47	0.52

Buy P

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
mean	2.96	2.96	2.83	2.89	2.84	2.87
median	3.25	2.75	2.75	2.75	2.75	2.75
standard deviation	0.60	0.52	0.71	0.60	0.64	0.66

Table 2: Valuations reported by individuals (in £; *n* = 175)

Sell $

	Round 1	Round 6
mean	5.39	4.14
median	5.25	3.75
standard deviation	2.16	1.99

Buy $

	Round 1	Round 6
mean	3.93	3.61
median	3.75	3.25
standard deviation	1.98	1.60

Sell P

	Round 1	Round 6
mean	3.49	3.15
median	3.25	2.75
standard deviation	1.18	1.08

Buy P

	Round 1	Round 6
mean	2.90	2.75
median	2.75	2.75
standard deviation	1.14	1.03

Table 3: Distributions of differences between asks and bids for individuals (*n* = 175)

	–2.50 or less	–2.00 or –1.50	–1.00 or –0.50	0.00	0.50 or 1.00	1.50 or 2.00	2.50 or more	mean difference
$: Round 1	3	9	22	20	54	34	33	1.46**
$: Round 6	5	13	48	20	50	22	17	0.53**
P: Round 1	2	5	39	34	71	14	10	0.60**
P: Round 6	0	5	47	51	52	14	6	0.40**

** denotes p < 0.01 for one-tailed t-test of hypothesis that difference between asks and bids is greater than zero.

Table 4: Distributions of differences between asks and bids for median traders (*n* = 33)

	–2.50 or less	–2.00 or –1.50	–1.00 or –0.50	0.00	0.50 or 1.00	1.50 or 2.00	2.50 or more	mean difference
$: Round 1	0	1	1	2	12	5	12	1.65**
$: Round 6	0	6	5	7	10	2	3	0.20
P: Round 1	0	0	6	10	12	4	1	0.38**
P: Round 6	0	0	7	14	10	1	1	0.15

** denotes p < 0.01 for one-tailed t-test of hypothesis that difference between asks and bids is greater than zero.

Table 5: Distributions of choices and valuations for individuals (*n* = 175)

	$ chosen			P chosen		
	V($)>*V* (P)	*V* ($)=*V* (P)	*V* ($)<*V* (P)	*V* ($)>*V* (P)	*V* ($)=*V* (P)	*V* ($)<*V* (P)
Selling: Round 1	84	5	**10**	**49**	13	14
Selling: Round 6	70	10	**19**	**42**	9	25
Buying: Round 1	76	13	**10**	**39**	10	27
Buying: Round 6	74	13	**12**	**41**	12	23

V($), *V*(P) denote valuations of S and P bets

Table 6: Distributions of choices and valuations for median traders (*n* = 33)

	$ chosen			P chosen		
	V($)>*V* (P)	*V* ($)=*V* (P)	*V* ($)<*V* (P)	*V* ($)>*V* (P)	*V* ($)=*V* (P)	*V* ($)<*V* (P)
Selling: Round 1	22	1	**0**	**10**	0	0
Selling: Round 6	20	1	**2**	**3**	2	5
Buying: Round 1	18	4	**1**	**5**	1	4
Buying: Round 6	19	2	**2**	**8**	2	0

V ($), *V* (P) denote valuations of $ and P bets

Table 7: Frequencies of classic and counter reversals for individuals with marginal bids or asks for both bets (*n* variable)

	Classic reversal	No reversal	Counter reversal	Total
Selling: Round 1	16	26	0	42
Selling: Round 6	17	36	5	58
Buying: Round 1	11	32	1	44
Buying: Round 6	13	46	3	62

Appendix: Identifying Preference Reversal at the Market Level

In this appendix, we assume that individuals act on preferences which satisfy the axioms of expected utility theory. Consider any individual. Let $u(.)$ be his utility function for wealth. Let w be his status quo level of wealth (to be treated as a constant) and let x be any small increment (to be treated as a variable). Then, using the first three terms of a Taylor expansion,

$$u(w + x) \approx u(w) + u'(w)x + u''(w)x^2/2. \tag{1}$$

Normalising by setting $u(w) = 0$ and $u'(w) = 1$, we have

$$u(w + x) \approx x - Rx^2/2, \tag{2}$$

where $R = -\,u''(w)/u'(w)$, i.e. the Arrow-Pratt measure of absolute risk aversion at w.

Now consider two lotteries L_1 and L_2, such that each L_k gives an increment of wealth x_k with probability p_k, otherwise giving an increment of zero. Assume $p_2x_2 > p_1x_1 > 0$ and $1 \geq p_1 > p_2 > 0$. Thus, L_2 has a higher expected value than L_1, but is unambiguously more risky. Treating the approximation (2) as an equality, the expected utility of L_k is $p_kx_k - Rp_kx_k^2/2$. Thus L_1 is strictly preferred to L_2 if and only if:

$$R < 2(p_2x_2 - p_1x_1)/(p_2x_2^2 - p_1x_1^2). \tag{3}$$

If the inequality in (3) is reversed, L_2 is strictly preferred to L_1; if it is replaced by an equality, the two lotteries are indifferent. We may interpret L_1 as a P bet and L_2 as a \$ bet. Thus, the value $R^* = 2(p_2x_2 - p_1x_1)/(p_2x_2^2 - p_1x_1^2)$ marks the boundary between a range of (lower) values of R at which \$ is preferred to P and a range of (higher) values at which P is preferred.

Next, consider the certainty equivalent value of any given lottery L_2 (which may now be interpreted either as a \$ bet or a P bet). This certainty equivalent is the value of x_1 such that, when $p_1 = 1$, $L_1 \sim L_2$. Using (3), it is straightforward to show that the certainty equivalent of L_2 falls as R increases.

Now consider a population of individuals 1, ..., n, indexed so that $R_1 \leq R_2 \leq \ldots \leq R_n$, where R_i is i's Arrow-Pratt measure of risk aversion at the relevant level of wealth. Let $k = (n + 1)/2$, i.e. k is the individual with the median degree of risk aversion. If, in a median price auction for a P or \$ bet, each individual's bid (or ask) is equal to his certainty equivalent value (that is, his underlying valuation according to the null hypothesis), the market price will be equal to *k*'s certainty equivalent. Thus, which of the two bets has the higher market price is determined by *k*'s preference between them. But, because *k* is the median individual and because each individual's preference between the bets is determined by his measure of risk aversion, whichever bet *k* prefers is chosen by the majority.

References

Ariely, Dan, George Loewenstein and Drazen Prelec (2003). "'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences". *Quarterly Journal of Economics* 118: 73-105.

Ballinger, T. Parker and Nathaniel Wilcox (1997). "Decisions, Errors and Heterogeneity". *Economic Journal* 107: 1090-1105.

Braga, Jacinto, Steven J. Humphrey and Chris Starmer (2006). "Market Experience Eliminates Some Anomalies – And Creates New Ones". CeDEx Discussion Paper 2006-19, University of Nottingham.

Butler, David and Graham Loomes (2007). "Imprecision as an Account of the Preference Reversal Phenomenon". *American Economic Review* 97: 277-297.

Camerer, Colin F. (1995). "Individual Decision Making." In *Handbook of Experimental Economics,* eds John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press.

Casey, J.T. (1991). "Reversal of the Preference Reversal Phenomenon." *Organizational Behavior and Human Decision Processes* (April) 48: 224-251.

Cherry, Todd L., Thomas D. Crocker and Jason F. Shogren (2003). "Rationality Spillovers." *Journal of Environmental Economics and Management* 45: 63-84.

Cherry, Todd L. and Jason F. Shogren (2007). "Rationality Crossovers." *Journal of Economic Psychology* 28: 261-277.

Chu, Yun-Peng and Ruey-Ling Chu (1990). "The Subsidence of Preference Reversals in Simplified and Marketlike Experiment Settings: A Note." *American Economic Review* 80: 902-911.

Coursey, Don L., John L. Hovis and William Schulze (1987). "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value". *Quarterly Journal of Economics* 102: 679-690.

Cox, James, C. and David M. Grether, D. (1996). "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory* **7**: 381-405.

Cubitt, Robin P., Alistair Munro and Chris Starmer (2004). "Testing Explanations of Preference Reversal". *Economic Journal* 114: 709-726.

Hanemann, Michael, W. (1999). "Neo-Classical Economic Theory and Contingent Valuation". In *Valuing Environmental Preferences*, eds Ian J. Bateman and Kenneth G. Willis. Oxford: Oxford University Press.

Hey, John and Chris Orme (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data". *Econometrica* 65: 581-598.

Knetsch, Jack L. and John A. Sinden (1984). "Willingness To Pay and Compensation Demanded: Evidence of an Unexpected Disparity in Measures of Value". *Quarterly Journal of Economics* 99: 507-21.

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

Knez, Marc and Vernon Smith (1987). "Hypothetical Valuations and Preference Reversals in the Context of Asset Trading." In *Laboratory Experimentation in Economics: Six Points of View*, ed. Alvin E. Roth (pp. 131-154). Cambridge: Cambridge University Press.

Lichtenstein, Sarah and Paul Slovic (1971). "Reversals of preferences between bids and choices in gambling decisions." *Journal of Experimental Psychology* 89: 46-55.

List, John A. (2002). "Preference Reversals of a Different Kind: The 'More is Less' Phenomenon." *American Economic Review* 92: 1636-1643.

List, John A. (2003). "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118: 41-71.

List, John A. and Jason F. Shogren (1999). "Price Information and Bidding Behavior in Repeated Second-Price Auctions". *American Journal of Agricultural Economics* 81: 942-49.

Loomes, Graham, Peter Moffatt and Robert Sugden (2002). "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice". *Journal of Risk and Uncertainty* 24: 103-130.

Loomes, Graham, Chris Starmer and Robert Sugden (2003). "Do Anomalies Disappear in Repeated Markets?" *Economic Journal* 113: C153-66.

Plott, Charles R. (1996). "Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis". In *The Rational Foundations of Economic Behaviour* eds Kenneth J. Arrow, Enrico Colombatto, Mark Perlman and Christian Schmidt. Basingstoke: Macmillan, pp. 225-250.

Seidl, Christian (2002). "Preference Reversal". *Journal of Economic Surveys* 6: 621-655.

Shogren, Jason F., Seung Y. Shin, Dermot J. Hayes and James B. Kliebenstein (1994). "Resolving Differences in Willingness to Pay and Willingness to Accept". *American Economic Review* 84: 255-270.

Shogren, Jason F., Sungwon Cho, Cannon Koo, John A. List, Changwon Park, Pablo Polo and Robert Wilhelmi (2001). "Auction Mechanisms and the Measurement of WTP and WTA". *Resource and Energy Economics* 23: 97-109.

Slovic, Paul and Sarah Lichtenstein (1968). "Relative Importance of Probabilities and Payoffs in Risk Taking". *Journal of Experimental Psychology* 78: 1-18.

Starmer, Chris (2000). "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk". *Journal of Economic Literature* 38: 332-382.

Sugden, Robert (1999). "Alternatives to the Neo-Classical Theory of Choice". In *Valuing Environmental Preferences*, eds Ian J. Bateman and Kenneth G. Willis. Oxford: Oxford University Press.

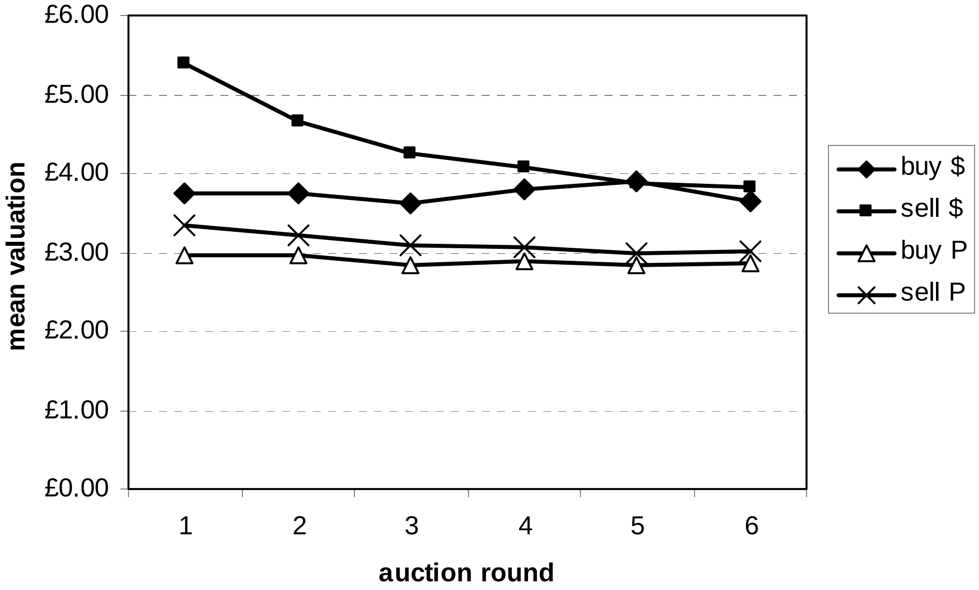
Sugden, Robert (2003). "Reference-Dependent Subjective Expected Utility". *Journal of Economic Theory* 111: 172-191.

Sugden, Robert (2004). "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences." *American Economic Review* 94: 1014-1033.

Tversky, Amos, Shmuel Sattath and Paul Slovic (1988). "Contingent Weighting in Judgement and Choice". *Psychological Review* 95: 371-384.

Tversky, Amos, Paul Slovic and Daniel Kahneman (1990). "The causes of preference reversal". *American Economic Review* 80: 204-17.

Figure 1: Valuations of median traders



Notes

[1] Hanemann (1999) has proposed a theoretical account of how such large disparities might occur within the framework of standard theory. But others have argued that, for plausible values of the relevant parameters, it is extremely difficult to generate the magnitude of disparity so often observed (e.g. Sugden 1999).

[2] Money-pumping at the individual level is possible only if arbitrageurs have monopoly access to particular consumers. Thus, economic competition tends to eliminate money-pumps, even if consumers have inconsistent preferences (Sugden, 2004).

[3] In an *k*th price Vickrey buying (respectively: selling) auction, each participant states a bid (ask). The $k – 1$ participants with the highest bids (lowest asks) buy (sell) at a price equal to the *k*th highest bid (*k*th lowest ask). In a median price auction, $k = (n + 1)/2$, where n is the number of participants, and n is odd. In a random *k*th price auction, the value of *k* is selected at random after bids (asks) have been stated.

[4] The repeated-market experiment of Ariely et al (2003, pp. 88-91) is a partial exception. In that experiment, participants reported WTA valuations for exposure to an annoying sound; after each market, participants with the lowest asks were paid the market price *and* experienced the sound.

[5] Sugden (2003) analyses a related mechanism by which loss aversion can contribute to preference reversals for WTA valuations.

[6] Indeed, if underbidding were the *only* source of deviations from underlying preferences, the broader interval of credible values for the $ bet might even result in an asymmetry in the opposite direction, with the majority of reversals involving choosing the $ bet but making a lower bid for it than for the P bet.

[7] We used different combinations of group sizes depending on the number of would-be participants who arrived at each session. By having two group sizes rather than one, we were able to use a larger proportion of arrivals as participants. We used odd-numbered groups so that the median price rule had a transparent definition.

[8] Instructions were read from a script, a copy of which is available from the authors.

[9] The results from this part of the experiment are reported in Loomes et al (2003).

[10] We used 0.19 and 0.81 as probabilities, rather than rounder numbers such as 0.2 and 0.8, to make the strategy of valuing bets at their expected values less salient to respondents.

[11] It is common (but not universal) for the bets used in preference reversal experiments to have this property.

[12] Compare the statements (i) 'The price is £5; you said you would buy if the price was £5; so you buy at a price of £5' and (ii) 'The price is £5; you said that the maximum price you would pay was £6; so you buy at a price of £5'. Statement (i) is more direct and so easier to understand.

[13] In our experience, people who are asked to report valuations of unfamiliar goods have a strong tendency to use round numbers, making the apparent precision of valuations elicited on finely graduated response scales largely spurious. Using 50p intervals keeps the valuation task for respondents relatively simple without much real sacrifice of precision. The lowest value in the set of £0.01 rather than zero because, for many people, the idea of 'buying' or 'selling' at a 'price' of zero is difficult to understand.

[14] For example, suppose the market price in a selling auction is £4.00. This implies that the participant who reported the median valuation was not willing to sell at £4.00, but was willing to sell at £4.50; according to our conventions, this is treated as a valuation of £4.25. Similarly, consider a buying auction in which the market price is £4.50. In this case, the participant who reported the median valuation was not willing to buy at £4.50, but was willing to buy at £4.00; this, too, is treated as a valuation of £4.25.

[15] The observation of such an asymmetry for selling valuations is in line with the mass of evidence from individual decision experiments. The great majority of preference reversal experiments have focused exclusively on selling valuations. Those that have used buying valuations have tended to produce less sharp results. It seems that the use of buying rather than selling valuations tends to reduce the frequency of classic preference reversals (Lichtenstein and Slovic, 1971; Knez andSmith, 1987) and may induce reversals in the opposite direction (Casey, 1991).

[16] This regularity in Hey and Orme's data was not reported in their 1994 paper, but was discovered in subsequent analysis. We are grateful to John Hey for this information.

[17] A preference reversal experiment reported by Butler and Loomes (2007) provides more direct evidence of the effect of repetition of choice tasks. Participants chose between the same P and $ bets on three separate occasions during the experimental session. There was no evidence of any tendency for P choices to become less frequent with repetition: the breakdown of P:$ choices for the 89 participants was 61:28, 68:21 and 64:25 for the first, second and third occasions respectively. An experiment reported by Braga, Humphrey and Starmer (2006) provides further evidence that the repetition of choice tasks has no systematic impact on the relative frequency of P and $ choices.

[18] If we consider the certainty of x as a limiting case of a P bet (i.e. in which the probability of winning tends to one), this behaviour is a limiting case of classic preference reversal.

[19] Cubitt, Munro and Starmer (2004) present several alternative hypotheses about the different heuristics evoked by choice tasks and by valuation tasks. In an experimental investigation, they find most support for the scale compatibility hypothesis.

[20] The experiment of Ariely et al (2003), in which anchoring effects are found to persist in repeated Vickrey auctions, provides further support for this claim.