The Economic Journal, **113** (*March*), C153–C166. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

DO ANOMALIES DISAPPEAR IN REPEATED MARKETS?*

Graham Loomes, Chris Starmer and Robert Sugden

There is some evidence that, as individuals participate in repeated markets, 'anomalies' tend to disappear. One interpretation is that individuals – particularly marginal traders – are learning to act on underlying preferences which satisfy standard assumptions. An alternative interpretation, the 'shaping' hypothesis, is that individuals' preferences are adjusting in response to cues given by market prices. The paper reports an experiment designed to discriminate between these hypotheses with particular reference to the disparity between willingness to pay and willingness to accept.

There is a large literature demonstrating the existence of 'anomalies' in individual choice behaviour; see Camerer (1995) and Starmer (2000). Taken at face value, such anomalies pose a major challenge to choice theorists, applied economists and policy analysts. However, much of this evidence comes from experiments involving one-off decisions in non-market settings, and some economists have questioned its economic significance by raising doubts about whether these anomalies will arise or persist in market environments. For example, Binmore (1994, 1999) argues that anomalous behaviour is economically significant only if it survives in an environment in which individuals repeatedly face the same decision problems, receive feedback on the outcomes of their decisions, and have adequate incentives.

In fact, there is some evidence that specific anomalies become less frequent in repeated experimental markets. Some of this evidence shows a particularly interesting feature: anomalies are eroded when individuals' preferences or valuations are elicited in repeated *markets*, but not when they are elicited by other mechanisms which include repetition, incentives and feedback (Cox and Grether, 1996; Shogren *et al.*, 2001).[1] These findings pose an important research question: how and why do repeated markets influence behaviour? Answering this question might help economists to define the domain in which standard assumptions about preferences are reliable. It might also inform the design of methods of eliciting individuals' preferences for use in policy analysis.

We examine two hypotheses which have been offered to explain the alleged tendency of repeated markets to eliminate anomalies, both of which imply that such markets lead at least some individuals to behave in accordance with underlying, context-independent 'true' preferences. We consider a third hypothesis about how stated values evolve in repeated markets which, if correct, suggests that much of the existing evidence of the erosion of anomalies might itself be an artefact of specific features of the experimental designs used. We then present the results of an experiment designed to discriminate between these hypotheses.

* We are grateful to three anonymous referees for comments on a previous version on the paper.

[1] Some other experiments using apparently similar repeated market designs have found that anomalies persist: for example, preference reversal decays in only one of the two experimental markets investigated by Cox and Grether, and Knetsch *et al.* (2001) find persistence of the WTP–WTA disparity.

1. The 'Discovery' of True Preferences: Two Hypotheses

The first hypothesis we consider is the *refining hypothesis*. This is that market experience has a *general* tendency to induce individuals to make decisions that increasingly accurately reflect their preferences. If preferences satisfy standard consistency requirements, and if anomalies result from errors, the refining hypothesis predicts a tendency for anomalies to become less frequent as market experience accumulates. This hypothesis does not specify the mechanisms which promote error reduction. It is simply an empirical conjecture in the spirit of Plott's (1996) 'discovered preference hypothesis'. Plott suggests that rationality is a 'process of discovery': when individuals face unfamiliar tasks, their behaviour can be influenced by various biases, but with incentives and practice, they arrive at 'considered choices' that reflect stable underlying preferences (Plott, 1996, p. 248). Much of the evidence cited by Plott concerns the remarkable tendency for various experimental market institutions to converge on the equilibrium predictions of standard economic models. Plott conjectures that this convergence is at least partly due to subjects discovering how best to satisfy their preferences.

The mechanisms of discovery are not fully specified by Plott, but neither are they totally mysterious. Like Plott and Binmore, we find it plausible to suppose that repetition, feedback and incentives might each play a role. Repetition allows subjects to become more familiar with decision tasks and the objects of choice; feedback allows subjects to experience the consequences of particular choices; and incentives provide a general motivation to attend carefully to tasks. If such factors tend to reduce error propensity, their operation could provide part of a more fleshed out version of the refining hypothesis.

Notice, however, that none of these factors is unique to a market context: it is possible to construct decision environments with repetition, feedback and incentives, but no market. As we have stated it, the refining hypothesis does not give any special significance to the market mechanism *per se*. The second hypothesis does.

Like the refining hypothesis, the *market discipline hypothesis* assumes that agents have stable underlying preferences, and that they may commit errors when attempting to act on those preferences within a market institution. However, the market discipline hypothesis distinguishes between two types of error: those which, *ex post*, are costly to the agent once the market outcome is known, and those which are not. The hypothesis is that agents adjust their behaviour to correct errors if and only if those errors have proved costly.

We consider the implications of this hypothesis for an auction in which each of n agents submits a bid to buy one unit of a good; the supply of the good is $k < n$ units. (A corresponding analysis applies for selling.) We assume that, for each agent j, the good has a (private) *value* v_j, indexed so that $v_1 \geq v_2 \geq \ldots \geq v_n$. There is *market-clearing equilibrium* in the Walrasian sense if the price p is in the interval $v_k \geq p \geq v_{k+1}$. The market institution in question is a generalisation of a Vickrey auction: the k agents submitting the highest bids buy at a price equal to the $(k + 1)$th highest bid. (The commonly used second price auction institution is the special case where $k = 1$.) Notice that if all agents bid their values, the price is equal to v_{k+1} and there is market-clearing equilibrium. It is well known that, for each participant in such an auction, it is a weakly dominant strategy to bid his value. However, since the market discipline hypothesis permits errors, it does not predict that bids are always equal to values.

Consider the cases in which the price is *not* market-clearing. First, suppose $p > v_k$. Then there must be at least one agent who bid more than his value and who, as a result, buys at a price greater than that value. This is a costly error. Alternatively, suppose $p < v_{k+1}$. Then there must be at least one agent who, as a result of bidding less than her value, misses an opportunity to buy at a price below that value. This too is a costly error. We take the market discipline hypothesis to imply that, in a repeated auction, an agent who makes a costly error in one round reacts by reducing the discrepancy between bid and value in the following round (if there is one). A simple version of this hypothesis is the partial adjustment model $b_{j,r+1} = b_{j,r} + \theta(v_j - b_{j,r})$, with $0 < \theta \leq 1$, where $b_{j,r}$ is the bid of agent j in round r and where that bid was a costly error. In this model, price converges to market-clearing equilibrium.

However, the hypothesis does *not* imply that *all* agents' bids converge to their values, because there can be a market-clearing equilibrium in which no costly errors are made but some bids are 'incorrect'. For example, there may be individuals who understate their values, but have no incentive to adjust because the market price is either above their value or below their bid. So, if the market discipline hypothesis is true, even indefinitely repeated auction mechanisms cannot be assumed to reveal the underlying preferences of *all* bidders. What *is* revealed, given sufficient repetition, is the market-clearing price that reflects the true preferences of *marginal* traders.

Notice that if the auction is represented as a game, any profile of bids which involves no costly errors is a Nash equilibrium. In game-theoretic terms, the market discipline hypothesis is an example of a myopic best-reply learning mechanism. That is, after making a costly error in a game, a player adjusts his behaviour in the direction of the best reply to others' current behaviour; this induces convergence to Nash equilibrium.

2. The Shaping Hypothesis

The refining and market discipline hypotheses both assume that each agent has true preferences that are independent of the mechanisms via which they are revealed or elicited. Both hypotheses assume that a repeated elicitation mechanism filters out some or all extraneous error and bias without affecting the preferences themselves. The assumption that preferences are 'mechanism independent' is obviously crucial if such hypotheses are to justify conventional economic theory as an explanation of behaviour in real repeated markets. But what if market experience alters or *shapes* preferences?

The *shaping hypothesis* is that, in repeated auctions in which prices have no information content, there is a tendency for agents to adjust their bids towards the price observed in the previous market period. Of course, if there is some element of common value in an auction, such an adjustment rule may be consistent with Bayesian updating of agents' beliefs about the value of the good for which they are bidding. But the shaping hypothesis applies to cases in which values are entirely private. The intuition behind the hypothesis is that, prior to her involvement in a specific market, an agent may not have well-articulated preferences waiting to be 'discovered'. Instead, values may be only partially formulated and/or imprecise, so that when confronted by an elicitation mechanism, responses are generated using heuristics in which market prices act as cues.

This hypothesis provides an alternative explanation for some experimental findings which have previously been interpreted as evidence in support of the refining or market discipline hypotheses.

Shogren *et al.* (2001) elicit willingness to pay (WTP) and willingness to accept (WTA) values in separate repeated second price auctions. The typical disparity between WTA and WTP is found in the early rounds of the auction but, with repetition, bids in the buying auctions rise and asks in selling auctions fall; after a few rounds, *average* WTP and WTA converge. This observation is consistent with both the refining and market discipline hypotheses, if each is supplemented by two additional assumptions. The first is that underlying preferences have the conventional property that WTP $\approx$ WTA. The second is that inexperienced subjects are subject to *strategic bias:* that is, they tend to bid low when buying and to ask high when selling. According to the refining hypothesis, we should expect such a bias to be eroded in all subjects through market experience. If the market discipline hypothesis is true, the bias may not be eliminated in *all* subjects, but those who experience costly errors will adjust their bids towards their values. Such adjustments will tend to reduce the understatement of values by buyers (thus increasing average WTP) and to reduce the overstatement of values by sellers (decreasing average WTA).

However, these data can also be explained by shaping. When subjects are *buying* in a second price auction, the market price is the second *highest* WTP; whereas in a selling auction, the market price is the second *lowest* WTA. So, when there are more than three traders in these auctions (as was the case in one of Shogren *et al.*'s experiments), the majority of bids will be below the market price, which shaping suggests will pull WTP up, while the majority of asks will be above the market price, which will generate a downward trend in WTA.

Cox and Grether (1996) report decay of preference reversal when lottery values are elicited in repeated auctions. In the classic preference reversal experiment, subjects confront 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. Subjects make straight choices between the two bets, and report a WTA valuation for each of them. The preference reversal anomaly is a widely observed tendency for subjects choosing the P bet to value the $ bet more highly. Cox and Grether elicit asks for P and $ bets in repeated second price auctions; after five asks have been elicited for each of the gambles, subjects make the straight choice. While the usual preference reversal pattern is observed when comparing first round bids with choices, there is no systematic pattern of reversals in the fifth round. This change occurs because asks for $ bets fall markedly across rounds.

A fall in WTA valuations for $ bets in this environment is consistent with all three hypotheses, for exactly the same reasons as are Shogren *et al.*'s findings. On all three hypotheses, of course, a declining trend of WTA valuations is also predicted for P bets. But because the probability of winning a P bet is high, reported valuations of a P bet tend to be concentrated in the narrow interval of 'reasonable' or 'credible' values bounded by the bet's expected value and the value of its prize. Thus, strategic bias is likely to induce greater overstatement of valuations for $ bets than for P bets. For the same reason, initial asks are more dispersed for $ bets than for P bets. Hence, shaping is likely to have a stronger impact on the distribution of valuations for $ bets.

3. A New Experimental Design

Our experiment was designed to discriminate between the three hypotheses in relation to disparities between WTA and WTP valuations of lotteries in repeated auctions. The design is built around two main innovations.

The first innovation is that we use *median price* auctions. In a median price buying auction, each participant reports the lowest price at which he is not willing to buy a specified lottery; the market price is the median of these 'just not willing to trade' bids; given this market price, all trades consistent with individuals' bids are then implemented. In a selling auction, each participant reports the highest price at which she is not willing to sell; the market price is the median of these asks; trades consistent with asks are then implemented. This market institution is a form of Vickrey auction which provides incentives for truthful revelation of values. By using median price auctions, we neutralise the effect, described in Section 2, by which shaping can induce convergence between WTA and WTP in second-price auctions. Thus, our design allows more demanding tests of the refining and market discipline hypotheses. Further, the median price rule ensures that *marginal* bids and asks in buying and selling auctions are comparable with one another: this is essential for a test of the market discipline hypothesis.[2]

The second innovation is a controlled test for shaping: we test whether stated valuations are influenced by the prices observed in previous market periods. To this end, we sought to construct an environment with two key features. First, the market should be such that agents are not sure of the value to them of the goods being traded. Second, the environment should be one in which different agents, bidding for identical goods in repeated markets, experience systematically different price feedback. Given that we required the market feedback to be a price genuinely produced by a freely operating market mechanism (and not fabricated by us as experimenters) the problem was how to engineer systematically different price feedback between the groups. Our solution was to construct auctions in which different participants bid to buy or sell *different* lotteries. In any given auction, each participant knows which lottery *he* is bidding to buy or sell; the market price is determined by the bids or asks of all participants; each participant who has indicated willingness to trade at the market price then does so, buying or selling his lottery at that price. By

[2] The *random nth price* rule used in some of the experiments reported by Shogren *et al.* (2001) is an alternative way of ensuring the comparability of marginal bids. Using the notation of Section 1, this form of Vickrey auction selects the value of k at random after bids have been submitted.

varying the distribution of lotteries involved in the auctions, we hoped to induce sufficient variation in market prices across auctions to allow a test of the shaping hypothesis. Notice that there is no common value component in these auctions. If, as the refining and market discipline hypotheses both assume, each participant has context-independent preferences between lotteries and money, then market prices contain no information that is relevant to the setting of bids or asks.

Although the main objective of the experiment was to elicit WTP and WTA valuations for lotteries, we also elicited valuations of *vouchers* using the same median price repeated auction mechanism. Each voucher had a fixed redemption value in money. As in the lottery auctions, different participants submitted bids or asks for different vouchers. In these treatments, the dominant strategy for each participant is to bid or ask according to the redemption value of her own voucher. The main purpose of this part of the experiment was to provide some indication of how well subjects understood the experimental environment: we wished to check that, in this respect, our experimental procedure was broadly comparable with those that other researchers have used. However, the voucher treatments also allow further tests of the refining and market discipline hypotheses. If subjects exhibit strategic bias, we can investigate whether disparities between bids or asks and redemption values reduce as markets are repeated.

We implemented this general design by assigning each subject at random first to a *trading group* and then, within each trading group, either to a larger *majority* subgroup or to a smaller *minority* subgroup. These groups and subgroups remained the same throughout the experiment. The members of each trading group participated together in a series of voucher and lottery markets.

The voucher treatments involved a *low value* voucher *VL* with a redemption value of £2.25 and a *high value* voucher *VH* with a redemption value of £4.25. The lottery treatments involved a *low probability* lottery *L*, offering a prize of £12 with probability 0.2 (and a 0.8 probability of winning nothing), and a *high probability* lottery *H*, offering a prize of £12 with probability 0.8. For each trading group, there were two voucher markets, one buying and one selling, and four lottery markets, two buying and two selling. Each of these markets was repeated six times in succession. The assignment of vouchers and lotteries to the two subgroups in these treatments is described in Table 1.

This structure elicits WTP and WTA valuations for each of the two lotteries in each of two different environments, a *Majority High* (henceforth *Maj H*) environment in which the majority subgroup is buying or selling *H* (markets 3 and 4) and

Table 1

Treatments Used in Experiment

Market	Type	Majority bid/ask for	Minority bid/ask for
1	buy vouchers	*VH*	*VL*
2	sell vouchers	*VL*	*VH*
3	buy lotteries	*H*	*L*
4	sell lotteries	*H*	*L*
5	buy lotteries	*L*	*H*
6	sell lotteries	*L*	*H*

a *Majority Low* (*Maj L*) environment in which this subgroup is buying or selling *L* (markets 5 and 6). We can expect the market price to be relatively high in the first environment and relatively low in the second. By comparing a given type of valuation (WTP or WTA) for given lotteries between *Maj H* and *Maj L* environments, we can test for shaping.

We test the refining hypothesis (while controlling for shaping effects) by holding the Majority condition constant and investigating whether differences between subjects' reported WTP and WTA valuations of given lotteries decline as markets are repeated. We also test the refining hypothesis in voucher markets by investigating whether disparities between bids (or asks) and redemption values decline with repetition.

We test the market discipline hypothesis (while controlling for shaping) by investigating whether, for a given environment, disparities between *median* WTP and WTA valuations within trading groups decline as markets are repeated. We also make analogous tests of that hypothesis in voucher markets comparing median valuations and redemption values across auction rounds.

4. Experimental Procedures

The data reported in this paper come from an experiment conducted at the University of East Anglia. A total of 175 subjects took part. Subjects were recruited from the general undergraduate population (i.e. not from any particular discipline or year of study). At the start of each experimental session, each subject was randomly assigned to a trading group either of five people, in which case three were assigned to the majority subgroup and two to the minority, or of seven people, in which case the majority was four and the minority three. In all, there were 33 trading groups; 104 subjects were assigned to majority subgroups and 71 to minority ones. In every session there were either two or three trading groups; individual subjects did not know who was in their own group. Throughout the experiment, subjects sat at individual computers with partitions between each person.

Each session began with an explanation of the procedures for the voucher auctions, structured around 'practice' auctions.[3] After taking part in two such practices (one buying and one selling, with different participants bidding for different vouchers), each trading group responded to the two voucher tasks in turn (markets 1 and 2 in Table 1), in random order, each market being repeated six times. Before these tasks, subjects were told: 'the value of the voucher [in each of the two voucher tasks] is different for different people in the room'. Then, the special features of lottery auctions were explained, and there were two practices of such auctions, followed by the four lottery tasks (markets 3 to 6 in Table 1), in random order, each being repeated six times.[4] In each of the practice auctions

[3] Instructions were read from a script, a copy of which can be obtained from the authors.

[4] In all, each subject took part in eight repeated lottery auctions in random order; in this paper we report data from only four. The other four auctions and two choice tasks, which tested different hypotheses, will be reported elsewhere. From the viewpoint of the subjects, the additional auctions were similar in form to those we report. The choice tasks were always the final tasks in the experiment and so subjects' exposure to these could not have affected the data reported here.

involving lotteries, all participants bought or sold the same lottery. Before going on to the 'real' lottery auctions, subjects were not told explicitly that different people might be bidding for different lotteries, but the instructions were phrased so as not to exclude this possibility. Subjects were told simply that in each task 'you will have the opportunity to buy a lottery ... or to sell [one]'.[5] In buying auctions, each subject was endowed with an amount of cash (which was always greater than the voucher redemption value or lottery prize). In selling auctions, each subject was endowed with the relevant voucher or lottery.

In each auction, bids or asks were elicited through an interactive computer program. We designed the elicitation procedure to be as simple and as transparent as possible and we developed and refined it through pilot experiments. To illustrate how it worked, consider a subject in a buying auction. The program asked a series of questions of the form: 'Would you be willing to pay £x?', adjusting the value of x according to previous answers. The series was structured so that non-monotonic responses were not possible. At the end of the sequence, the computer summarised the implications of the subject'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 subject had said she would pay and x'' (such that $x'' > x'$) was the smallest amount she had said she would not pay. The computer then asked the subject to confirm that she was happy with this statement. If the subject confirmed, x'' was recorded as the subject's 'just not willing to trade' value; if the subject did not confirm, the elicitation procedure recommenced.

The set of possible values of x was {£0.01, £0.50, £1.00, ..., £4.50, £5.00} for vouchers and {£0.01, £0.50, £1.00, ..., £11.50, £12.00} for lotteries.[6] In reporting results, we use a scale on which 0 denotes a valuation of less than £0.01, 1 a valuation between £0.01 and £0.50, 2 a valuation between £0.50 and £1.00, and so on in increments of £0.50. For vouchers, 11 denotes a valuation greater than £5.00; for lotteries, 25 denotes a valuation greater than £12.00. (For example, a subject who is willing to buy a lottery at £6.00 but not at £6.50, or who is willing to sell at £6.50 but not at £6.00, is reported as recording a valuation of 13.)

After each round of each auction, each subject was told the market price for that round and its trading implications for her (i.e. whether she bought or sold the voucher or lottery, and if so, how much she would pay or receive). Lotteries were not played out at this stage. Each subject knew from the outset that one round of one auction would be selected at the end of the experiment, and that whatever decisions she had made in the selected round would be implemented. If the round selected was a selling round and if the subject had sold her voucher or lottery during that round, she was paid whatever had been the market price. If she had kept her endowment, she was paid the redemption value if it was a voucher or, if it

[5] The intention was that this procedure, while not involving deception, would suggest to subjects that they might all be bidding for the same lottery. Such beliefs on the part of subjects could be expected to increase the strength of shaping effects. Notice that, if a subject has context-independent preferences between lotteries and money, her beliefs about what other subjects are bidding for are irrelevant for her own bidding strategy.

[6] £0.01 was used instead of zero because, in our pilot studies, we found that subjects had difficulty with the concept of paying (or accepting) zero.

was a lottery, the lottery was played out and the subject received the outcome. If the round selected was a buying round, in addition to the payoff from any lottery or voucher that may have been purchased, the subject was paid any remaining cash endowment (net of the purchase price if she had bought). Each experimental session lasted about 60 minutes, and the average payment per subject was £7.06.

5. Results: Voucher Auctions

Table 2 summarises the results of the voucher auctions. The first four rows of the Table report the distributions of bids and asks of the 175 subjects in the first and last rounds of the two auctions. The data are reported as deviations between the valuation of each voucher implied by a subject's bid or ask (on the scale from 0 to 11) and its redemption value (i.e. 9 for the high value voucher, 5 for the low value voucher). The remaining four rows of the Table report the corresponding distributions of median bids and asks in the 33 trading groups. In this case, the data are reported as deviations between the valuations implied by median bids or asks and median redemption values (i.e. 9 for buy auctions, where the majority were trading high value vouchers, and 5 for sell auctions, where the majority were trading low value vouchers).

The majority of subjects made correct bids or asks, and many of the differences between actual and correct responses were relatively small. We take this finding as reassurance that most subjects were responding coherently and sensibly within our auction mechanism.[7] However, among those who do not make the correct bid or

Table 2

Distributions of Deviations in Voucher Auctions

	Recorded bid/ask minus redemption value:								
	<−2	−2	−1	0	1	2	>2	Mean	St.dev
By subjects ($n = 175$):									
buy: round 1	18	13	28	92	12	8	4	−0.53**	1.82
buy: round 6	25	16	22	95	6	7	4	−0.74**	1.89
sell: round 1	3	3	14	112	16	12	15	0.38**	1.48
sell: round 6	6	3	14	107	22	15	8	0.25*	1.42
By trading groups ($n = 33$):									
buy: round 1	4	5	6	12	6	0	0	−0.70**	1.36
buy: round 6	9	5	4	6	9	0	0	−1.12**	1.93
sell: round 1	0	0	0	14	0	7	12	1.70**	1.63
sell: round 6	0	0	0	16	6	6	5	1.09**	1.33

In all cases, the null hypothesis is that mean deviations are not significantly different from zero. For buying auctions, the alternative hypothesis is that there is underbidding (deviations < 0); while for selling auctions, the alternative hypothesis is that there is overasking (deviations > 0). *(**) denotes rejection of the null in favour of the alternative at the 5% (1%) level of significance, using a one-tailed z-test.

[7] The frequency of incorrect bids is in line with the findings of other auction experiments using induced values. In such experiments, deviations of bids from values are not uncommon among non-marginal traders (Miller and Plott, 1985; Franciosi *et al.*, 1993).

ask, there are asymmetries in the deviations: there is a clear and statistically significant tendency for underbidding in buying auctions and for overasking in selling auctions. The most obvious interpretation of these asymmetries is strategic bias. There seem to be no substantial changes in the distributions of individuals' deviations over the six rounds. The round 1 and round 6 distributions are remarkably similar: indeed, the degree of underbidding has actually increased slightly; and although the degree of overasking has reduced a little, it is still strongly significant in round 6 ($p = 0.011$). These results provide no support for the refining hypothesis.

To investigate whether bidding behaviour is consistent with the market discipline hypothesis, we need to look at deviations between *median* reported valuations and *median* redemption values within trading groups. The median valuation is equal to the redemption value in fewer than half of the markets, and the distributions of deviations are again asymmetric, with highly significant evidence of underbidding and overasking and little evidence of any strong tendency for the asymmetries to be eliminated. There is little or no support here for the market discipline hypothesis.

6. Results: Lottery Auctions

The results of the lottery auctions are summarised in Tables 3, 4 and 5. Table 3 reports the mean WTP and WTA valuations implied by subjects' bids and asks for the two lotteries, in the *Maj H* and *Maj L* environments, in the first and last auction rounds. Table 4 reports averages, across the 33 trading groups, of the *median* WTP and WTA valuations of the two lotteries in the two environments, in the first and last rounds. Table 5, which is the lottery auction analogue of Table 2, reports distributions of within-subject and within-market WTA-WTP disparities.

We can test the refining hypothesis by investigating whether, for each lottery and for each market environment, disparities between mean WTA and WTP valuations decline over the six rounds. The last column of Table 3 shows that, in all four cases, WTA is greater than WTP in both the first and the last rounds, but the disparity is less in round 6 than in round 1. The disparity is statistically significant

Table 3

Mean Valuations in Lottery Auctions

Lottery traded	*Maj*	Round	*n*	WTP: mean (s.d.)	WTA: mean (s.d.)	Mean WTA minus mean WTP
L	*L*	1	104	7.03 (4.11)	8.72 (4.65)	1.69**
L	*L*	6	104	7.26 (3.81)	8.10 (4.11)	0.84*
L	*H*	1	71	7.42 (3.61)	9.24 (4.17)	1.82**
L	*H*	6	71	8.41 (3.86)	9.41 (3.59)	1.00
H	*L*	1	71	14.15 (5.08)	16.89 (3.56)	2.74**
H	*L*	6	71	12.06 (4.63)	14.24 (5.09)	2.18**
H	*H*	1	104	14.12 (4.98)	15.06 (4.92)	0.94
H	*H*	6	104	12.48 (4.55)	13.29 (4.80)	0.81

*(**) denotes that WTA minus WTP is significantly greater than zero at the 5% (1%) level, using a one-tailed z-test.

Table 4

Median Valuations in Lottery Auctions

Market with *Maj* trading	Round	*n*	WTP: mean (s.d.) of medians	WTA: mean (s.d.) of medians	WTA minus WTP
L	1	33	8.70 (3.22)	10.85 (4.17)	2.15**
L	6	33	9.12 (2.92)	9.00 (3.20)	−0.12
H	1	33	11.39 (2.52)	12.85 (3.16)	1.46*
H	6	33	11.30 (2.64)	11.00 (3.05)	−0.30

* (**) denotes that WTA minus WTP is significantly greater than zero at the 5% (1%) level, using a one-tailed z-test.

Table 5

Distributions of WTA-WTP Disparities in Lottery Auctions

	WTA minus WTP:							
	<−6	−6 to −4	−3 to −1	0	1 to 3	4 to 6	>6	st.dev
By subjects ($n = 175$):								
Maj L, round 1	4	6	50	15	47	30	23	5.70
Maj L, round 6	10	13	44	21	42	21	24	5.72
Maj H, round 1	12	9	41	21	39	29	24	5.76
Maj H, round 6	9	22	32	28	42	19	23	5.42
By trading groups ($n = 33$):								
Maj L, round 1	1	2	4	5	9	9	3	4.31
Maj L, round 6	1	5	7	5	10	4	1	3.59
Maj H, round 1	0	5	4	4	10	7	3	3.95
Maj H, round 6	1	4	11	4	9	2	2	3.68

in three out of four cases in round 1, and in two out of four cases in round 6. These results give qualified support to the refining hypothesis insofar as there is *some* overall tendency to erode the disparity between WTA and WTP.

To test the market discipline hypothesis, we consider disparities between the WTA and WTP valuations of *marginal* traders. Table 4 shows that WTA exceeds WTP for marginal traders in round 1 to a degree that is significant at the 1% level in the *Maj L* case and at the 5% level in the *Maj H* case. However, the disparity has entirely disappeared by round 6. This result would appear to provide clear support for the market discipline hypothesis.

The results presented in the previous two paragraphs are based on *averages* of the bids and asks of different individuals and on *averages* of buying and selling prices generated for different trading groups. While those results suggest that market experience erodes the *systematic* component of the disparity between WTA and WTP, that does not necessarily imply that market experience induces *convergence* between WTA and WTP, either for individual agents or for individual markets. To investigate the extent of convergence, we need to consider the data in Table 5.

The most striking feature of these data is the high degree of dispersion in *all* the distributions of differences between WTA and WTP. If we take the standard deviations as a summary measure of dispersion, there is *some* reduction between round 1 and round 6, but the reduction is modest. So if it is assumed that individuals have well-defined preferences which imply WTA $\approx$ WTP, it cannot be claimed that the repeated auction mechanism succeeds in accurately eliciting those preferences at the level of the individual – whether or not that individual is a marginal trader. At best, the mechanism is eliciting those preferences with a high degree of noise. An alternative interpretation is that individuals are uncertain about what their preferences are, and that market experience does not eliminate that uncertainty.

We now consider the shaping hypothesis. The most straightforward test of this hypothesis is to compare, for each lottery, the valuations reported in round 6 in the *Maj H* and *Maj L* environments. These are between-subject tests: we are comparing the bids or asks made by respondents randomly allocated to two subgroups, between which the only systematic difference lies in the feedback they have been given about market prices. The null hypothesis is that the Majority condition makes no systematic difference to the round 6 valuations. The alternative hypothesis implied by the shaping hypothesis is that round 6 valuations are higher in the *Maj H* environment.

Using the data in Table 3, four such comparisons can be made. For valuations of *L* (rows 2 and 4 of the Table), we see that mean WTP was 8.41 in the *Maj H* environment, compared with 7.26 under *Maj L*; while for WTA, the corresponding comparison was 9.41 as against 8.10. For valuations of *H* (rows 6 and 8), the WTP means were 12.48 under *Maj H* as compared with 12.06 under *Maj L*, while for WTA the corresponding figures were 13.29 and 14.24. Thus in three of the four cases (WTP for *L*, WTA for *L*, and WTP for *H*), the difference between the *Maj H* and *Maj L* mean valuations is in the direction predicted by the shaping hypothesis, and in two of these cases (WTP for*L* and WTA for *L*) the difference is significant at the 5% level (using a one-tailed z-test).

One possible limitation of the test reported in the previous paragraph is that it takes no account of differences between the subgroups that may have been present (albeit by chance) in round 1. An alternative strategy is to adjust for such differences by computing *changes* in subjects' bids and asks between round 1 and round 6, and then to compare these changes between the *Maj H* and *Maj L* conditions. The null hypothesis of no shaping is that the Majority condition has no systematic effect on the way bids or asks for given lotteries change as subjects accumulate experience. Shaping implies that, relative to the *Maj L* feedback, the *Maj H* feedback exerts a net upward pull on bids and asks for both lotteries; thus, the net increase in bids and asks for each lottery between round 1 and round 6 should be greater in the *Maj H* condition.

Table 6 shows the relevant comparisons. In all four cases the differences between mean changes are in the direction consistent with shaping. In three of these cases, the difference is significant at the 10% level.

At first sight, the results of these two sets of tests might seem to suggest that shaping has an effect, but only a rather modest one. However, when assessing the weight of this evidence, it is important to remember that, if shaping effects exist, their magnitude will depend on the magnitude of the differences in the price

Table 6

Changes in Bids/Asks, Round 1 to Round 6

Lottery traded	*Maj*	Bid/Ask	*n*	Mean change	Standard deviation	Significant difference?
L	*H*	Bid	71	+0.99	3.74	
L	*L*	Bid	104	+0.23	2.89	10%
L	*H*	Ask	71	+0.17	3.78	
L	*L*	Ask	104	−0.625	3.71	10%
H	*H*	Bid	104	−1.63	4.75	
H	*L*	Bid	71	−2.10	4.35	No
H	*H*	Ask	104	−1.77	3.69	
H	*L*	Ask	71	−2.65	4.39	10%

feedback which generate them. In our experiment, the relevant difference is best measured by the differences between the market prices generated in the first round in the *Maj H* and *Maj L* conditions. On average, the *Maj H* prices were 25% greater than *Maj L* prices (see Table 4).

Why was the difference between prices in the two conditions so small, when *H* offered four times as much chance of the same £12 payoff as L? In order to be able to compare like with like when examining the relationship between WTA and WTP, we set market prices at the valuation of the median participant. Typically, therefore, in a *Maj H* auction the market price will represent the *lowest* bid or ask for *H*, while in a *Maj L* subgroup, it will represent the *highest* bid or ask for *L*. This tends to compress the difference between the prices fed back to subjects in our median price design. When auctions involve a large number of participants and use a second price mechanism (as in several of the studies cited earlier), the difference between the second highest and the second lowest valuation – the difference that generates shaping effects – may be much greater than this.

7. Conclusions

Our experiment was motivated by two related conjectures. The first was that bids and asks in repeated markets might be influenced by shaping effects. The second was that existing evidence suggesting that disparities between WTA and WTP are eroded by market experience might be an artefact of experimental designs in which valuations are contaminated by shaping effects.

With respect to the second of these conjectures, our results suggest that market experience *does* tend to erode whatever causal factors generate the tendency for WTA to be systematically greater than WTP. Our findings are not sharp enough to allow us to conclude whether the mechanism by which the disparity is eroded is that of the refining hypothesis or that of the market discipline hypothesis. Since we do not find any corresponding tendency for market experience to reduce underbidding and overasking for vouchers with fixed redemption values, it seems likely that the factor being eroded is something other than, or additional to, strategic bias. In psychological terms, one possible explanation is that loss aversion makes people reluctant to accept exchanges which require them to give up an initial endowment and exchange it for something different; but by participating in markets in which endowments are routinely sold, they become more familiar with, and so less averse to, the idea of selling. However, our results also suggest that, even after repeated trading, individuals' valuations of given lotteries remain subject to a high degree of stochastic variation, arguably reflecting many subjects' continuing uncertainty about what these lotteries are really worth to them. Thus, the disappearance of the systematic component of the WTA–WTP disparity should not be interpreted as convergence to precise and stable 'true' preferences.

With respect to the first conjecture, our results suggest that systematic shaping effects *do* occur. If behaviour in markets is indeed influenced by shaping, the validity of repeated market mechanisms as means of eliciting individuals' preferences is called into question. If such mechanisms are to be used for this purpose, they need to be designed in the light of an understanding of the dynamics of shaping. It is of course possible that shaping – like loss aversion on the interpretation we offered in the previous paragraph – is itself a bias which market experience eventually eliminates. But our results suggest grounds for scepticism on this score. Our conjecture is that shaping is associated with preference imprecision: the less sure a person is about what his preferences 'really' are, the more susceptible he is to external cues such as information about market prices. If that is right, we should expect any erosion of shaping effects to be associated with a reduction in the stochastic component of individuals' preferences; but our results give little support to the idea that preference imprecision declines with market experience. Clearly, however, these are issues to be resolved by further empirical research.

University of East Anglia
University of Nottingham
University of East Anglia

References

Binmore, K. (1994). *Playing Fair*, Cambridge, MA: MIT Press.

Binmore, K. (1999). 'Why experiment in economics?', ECONOMIC JOURNAL, vol. 109, pp. F16–24.

Camerer, C. F. (1995). 'Individual decision making', in J. Kagel and A. E. Roth, eds., *Handbook of Experimental Economics*, pp. 587–703, Princeton: Princeton University Press.

Cox, J. C. and Grether, D. M. (1996). 'The preference reversal phenomenon: response mode, markets and incentives', *Economic Theory*, vol. 7, pp. 381–405.

Franciosi, R., Isaac, R. M., Pingry, D. and Reynolds, S. (1993). 'An experimental investigation of the Hahn-Noll revenue neutral auction for emissions licenses', *Journal of Environmental Economics and Management*, vol. 24, pp. 679–90.

Knetsch, J., Tang, F. F. and Thaler, R. (2001). 'The endowment effect and repeated market trials: is the Vickrey auction demand revealing?', *Experimental Economics*, vol. 4, pp. 257–69.

Miller, G. and Plott, C. (1985). 'Revenue generating properties of sealed-bid auctions: an experimental analysis of one-price and discriminative auctions', in V. Smith, ed., *Research in Experimental Economics*, vol. 3, pp. 159–82, Greenwich, Connecticut: JAI Press.

Plott, C. (1996). 'Rational individual behavior in markets and social choice processes: the discovered preference hypothesis', in K. Arrow, E. Colombatto, M. Perleman and C. Schmidt, eds., *Rational Foundations of Economic Behavior*, pp. 225–50, London: Macmillan.

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*, vol. 23, pp. 97–109.

Starmer, C. (2000). 'Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk', *Journal of Economic Literature*, vol. 38, pp. 332–82.