Preference Reversal

Jacinto Braga

University of Nottingham, June 2003

Reversal of preference between values and choices in gambling decisions[1] has been observed in several experiments: quite often subjects choose the gamble they value less highly. The phenomenon, as it has been called, seems to imply intransitive preferences or violation of procedure invariance. Any of these would be a major departure from the traditional theory of rational choice. The phenomenon has motivated substantial empirical and theoretical research (see Seidl 2002 for an extensive review). This chapter presents an overview of that research, and identifies open questions, some of which provide the topics for the subsequent chapters.

1. Preference Reversal and the Theory of Rational Choice

Consider the following pair of bets:

P = (-\$1, $\frac{1}{36}$; \$4, $\frac{35}{36}$), and $\$$ = (-\$1.5, $\frac{25}{36}$; \$16, $\frac{9}{36}$).

The notation is read thus: for example, the *P* bet offers one chance out of 36 of losing \$1, and 35 chances out of 36 of winning \$4. This pair of bets was actually used in several experiments and in the theoretical literature. In a typical experiment all pairs are made up of a *P* bet and a *\$* bet with varying probabilities and amounts to win, but sharing the following feature: the *P* bet offers a higher probability of winning than the *\$* bet, whereas the *\$* bet offers more money to win.

In a typical experiment subjects see one pair at a time, and are asked to name the bet they would rather play. Then they face one bet at a time, and are asked to

[1] Rephrasing of the title of Lichtenstein and Slovic (1971).

place their monetary value on it. Economists in particular like their experiments to be *incentive compatible*, that is they like the experimental design to make it the subjects' best interests to reveal their true preferences. In choice tasks this can be easily achieved by letting subjects play the bet they choose. In valuation tasks incentive compatibility has often been sought by using a procedure devised by Becker, De Groot, and Marshak (1964), known since then by their names, or by BDM procedure for short. It works as a second-price selling auction in which each subject competes with a device that makes random bids. Subjects are told to imagine they own the right to play a bet, and are offered the possibility of selling it. They are asked to state the minimum price they would sell the lottery for. Call this price their valuation. Then the experimenter makes a random offer for the lottery. If this offer price is equal to or higher than the subjects' valuation they sell the lottery for the offer price; otherwise they keep and play the lottery.

This procedure gives subjects an incentive to state as their valuation of a bet its certainty equivalent. Suppose someone's certainty equivalent of a bet is £3. Then if the offer price was any higher than £3, they would prefer to receive the offer price rather than play the lottery. Therefore they should value the bet at no more than £3. If the offer price was £2.99 or less they would rather play the lottery than receive the offer price. So they should value the bet at no less than £3. Therefore if they value it at £3 they will guarantee themselves whichever they prefer, the offer price or the lottery. In fact as valuations and offer prices must be in pounds and whole pence, subjects whose certainty equivalent is £3 would also guarantee themselves their preferred outcome if their valuation is £3.01.[2]

In experiments run along these lines subjects quite often choose the P bet, but value the $\$$ bet more highly than the P bet. In these cases, denoting v_P and $v_\$$ the valuations of the P and $\$$ bets, we have

$$\$ \sim v_\$ \succ v_P \sim P.$$

That is, he is indifferent between the $\$$ bet and $v_\$$, and between v_P and the P bet. Under the natural assumption of monoticity he prefers the higher $v_\$$ to the lower v_P. So assuming that the subject's preferences are transitive we conclude that he prefers the $\$$ bet to the P bet. But from his choice we conclude the reverse.

[2] Actually all this reasoning relies on the independence axiom of preferences over lotteries, but apparently that went unnoticed for over a decade, and we will ignore it for a while as well.

So the following decision patterns,

$$P \succ_c \$ \quad \text{and} \quad v_{\$} > v_P, \tag{1}$$

$$\$ \succ_c P \quad \text{and} \quad v_P > v_{\$}, \tag{2}$$

where $\succ_c$ means "is chosen over," are what has been called preference reversals. These preference reversals seem to go counter any theory of rational choice. And they have been proved robust: they have been observed in several experiments with different designs, and run by different people.

Before we proceed it is useful to clarify what is meant in this thesis by a *theory of rational choice* or *optimising behaviour*, phrases that will be used interchangeably. Such a theory is defined by the following two assumptions: individuals evaluate the consequences of their possible courses of action, and opt for the course of action leading to their preferred consequence; and the evaluation is independent of the process mediating action and consequence.[3] This independence has been called *procedure invariance*, and is a crucial feature of any theory of rational choice.

Standard economic theory of choice has assumed procedure invariance as a matter of course, and is therefore a theory of rational choice according to the definition adopted here. The reason to make procedure invariance explicit here is that we will see below theories of decision that do not assume it. Thus, in this thesis, procedure invariance will be the distinguishing feature of a theory of rational choice.

Standard economic theory, in addition to procedure invariance, typically assumes that the evaluation of consequences is made according to a weak preference relation, $\succeq$, that is complete, reflexive, transitive, and monotonic. Strict preference will be denoted by $\succ$, and indifference, by ~. The most popular theory of rational choice under uncertainty, expected utility, additionally assumes that preferences between two lotteries are independent of common parts.

In light of these clarifications what is the challenge posed by preference reversal to the standard economic theory of rational choice? Consider our pair of *P* and *$* bets. Suppose that valuations are elicited with the BDM procedure, and that an

[3] Individuals are naturally entitled to have preferences over aspects of those mediating processes. In that case such aspects should be included in the consequences to be evaluated.

individual values the *$* bet at $5, and the *P* bet at $3. Assuming that the individual evaluated correctly the consequences of these and the alternative valuations, he knows that if a counter offer of, say, $4.99 is made for the *$* bet, he will keep and play the *$* bet rather than receive $4.99. Thus *$* $\succeq$ $4.99. Assuming monotonicity, $4.99 $\succ$ $3. Reasoning in the same way, $3 $\succeq$ *P*. Assuming transitivity, *$* $\succ$ *P*. And if the individual chooses *P* over *$*? Possible explanations would be non-monotonic or non-transitive preferences. We will see below that another explanation is the violation of the independence axiom of expected utility. Yet another explanation would be that the evaluation of consequences in choices differs from the evaluation of the same consequences in valuations, that is, a violation of procedure invariance. Violation of procedure invariance would be a rejection of any theory of rational choice. The other explanations would imply at least the rejection of the most popular economic theory of choice under uncertainty.

2. How it Began

Preference reversals were first reported by psychologists Sarah Lichtenstein and Paul Slovic in 1971, but they had predicted them earlier. They had observed in an experiment, reported in Slovic and Lichtenstein (1968), that choices between bets correlated mostly with probabilities, whereas monetary valuations correlated mostly with the amounts to win. This led them to think that if they constructed pairs of bets with the features of the pair above, that is, consisting of a *P* bet with a high probability of winning a modest amount, and a *$* bet with a small probability of winning a largish amount, some subjects would choose the *P* bet but place a higher value on the *$* bet, thus exhibiting pattern (1) of preference reversals. Their intuition turned out to be right: in three experiments consisting of choice and pricing tasks similar to those described above they observed plenty of instances of pattern (1), but few of pattern (2). Following Tversky et al (1990), we will refer to pattern (1) as *standard reversal*, and pattern (2) as *non-standard reversal*.

In experiments 1 and 2 the payoffs of the bets were imaginary, and subjects were paid by the hour. In one of these experiments subjects were asked to state the minimum price they would sell the bet for (bid to sell); in the other they stated the maximum price they would pay for the bet (bid to buy). In the third experiment subjects either played or sold the bets for real, the sale being according to the BDM procedure. But the payoffs of the bets were expressed in points that were at the end of the experiment converted into dollars, so that the actual payment a subject might receive ranged from 80 cents to 8 dollars.

The following table summarises the results of the experiments.

Table 1: Response Patterns* (%), in Lichtenstein and Slovic (1971)

Experiment	Preference Reversals		Consistent Decisions†
	Standard	Non-standard	
Imaginary payoffs			
Bids to sell	42.5	3.1	54.4
Bids to buy	27.1	12.7	60.2
Real payoffs, bids to sell	32.1	4.8	63.1

* For example, the 42.5 on the top left means that of all choices made by all subjects in that experiment 42.5% were cases in which subjects chose *P* and valued *$* more highly than *P*.

† Choosing *P* and valuing *P* more highly than *$*, or choosing *$* and valuing *$* more highly than *P*.

Source: Lichtenstein and Slovic (1971).

Lichtenstein and Slovic (1973) replicated these results in an experiment with gamblers betting their own money in a casino in Las Vegas.

Harold Lindman (1971) also found evidence of preference reversals in a different experimental setting. Although he does not report the frequency of reversals he found that the gamble with the highest average price across subjects tended not to be the most often chosen one. However when subjects performed the experiment for the sixth time prices had moved more in accordance to choices.

Lichtenstein and Slovic interpreted these results as a violation of procedure invariance, and thus as contradicting the economic theory of rational choice. Economic theory assumes that individuals have well defined preferences, and that revealed preferences are invariant to the procedure used to elicit them. Their contrasting view is that the preferences people reveal in their decisions depend on the way they process information, and this in turn depends on the elicitation procedure, or response mode, for instance choices and valuations.

They argue that when asked to value a bet, as the response must be an amount of money, subjects anchor on the amount to win and adjust it downwards, taking account of the probabilities. This adjustment would be insufficient because translating probabilities into money amounts requires effort. Therefore valuations are mostly influenced by the amount to win. This became known as the *compatibility* *hypothesis*: the idea that the weight an attribute is given in a response is greater the more compatible the attribute is with the response mode. Thus if the response mode is valuation the attribute of the bet most compatible with it is the amount to win. This should lead to an overvaluation of *$* bets, because, as their winning probabilities are small, the downward adjustment should be quite sizeable if probabilities were taken full account of. Overvaluation of *P* bets should be far more modest, as their high winning probabilities require only small downward adjustments anyway.

In choices none of the dimensions of the bet, probabilities and amounts to win and lose, is any more compatible than the others with the response mode. Therefore the compatibility hypothesis does not predict choices to be mostly influenced by any of the dimensions of the bet. This combined with the tendency to overvalue the *$* bets should lead to standard reversals.

3. Economists Meet Preference Reversal

The first published reaction from economic quarters to this challenge to utility theory came from Grether and Plott (1979). They conducted several experiments "designed to discredit the psychologists' work as applied to economics" (p. 623), but ended up concluding that in their experiments "the preference reversal phenomenon which is inconsistent with the traditional statement of preference theory remains" (p. 634).

Grether and Plott's (1979) initial reaction to the inconsistencies observed in past experiments was that they could have been caused by a number of factors other than the violation of procedure invariance suggested by psychologists. Their experiments were designed to control for those alternative explanations, but the phenomenon persisted, leading the authors to reject the alternative explanations and accept the violation of procedure invariance. Grether and Plott's study was followed by another two, one by Pommerehne, Schneider and Zweifel (1982), and the other by Robert Reilly (1982). These authors were somewhat sceptical about Grether and Plott's results. However, although they observed less preference reversal in some of their experiments, the general conclusion was that the preference reversal phenomenon is robust.

One central question in all these three papers was whether subjects were sufficiently motivated to make decisions that really reflected their preferences.

Most of the previous experiments involved only gambles with imaginary payoffs, so, Grether and Plott argued, subjects' decisions were meaningless to economics because subjects lacked incentives to truthfully reveal their preferences. Related to this argument is the idea that making careful decisions requires mental effort. In the absence of strong incentives to compensate them for incurring these decision costs, subjects might resort to simpler but inaccurate decision rules, maybe of the type put forth by psychologists in their compatibility hypothesis. Grether and Plott also dismissed Lichtenstein and Slovic's (1971) experiment with actual gambles on these grounds, because although gambles were played for real their payoffs were expressed in points, and subjects were not told the exact conversion of points into money until the end of the experiment.

Also related to the incentives question are income effects. In experiments in which payoffs were real, and thus were not liable to the lack of incentives criticism, subjects played or sold all gambles. Therefore their income changed during the experiment, or, in case the gambles were all played at the end of the experiment, the expected income changed. Therefore subjects' attitude towards risk might have changed, which might have accounted for apparent preference reversals.

Grether and Plott dealt with these incentive-related questions in their experiment 1. It resembled Lichtenstein and Slovic's (1971) experiment with incentives. The same pairs of bets were used, and subjects undertook the choice and the valuation tasks, the latter using the BDM procedure to elicit minimum selling prices. To address the incentive problem two groups of subjects faced different incentive schemes: subjects in one group were paid a fixed seven dollars; in the other group they played or sold bets for real. To avoid the effects of changing income in the latter group only one task, a choice or a valuation, was selected at random at the end of the experiment to be played for real. This practice came to be known as the *random lottery selection*, and found widespread use in experiments thereafter. All studies reviewed here used it unless stated otherwise.

The conclusion to draw from Grether and Plott's experiment is that neither the lack of incentives nor income effects explain preference reversals: in both groups the familiar pattern emerged, there were lots of standard reversals and few non-standard ones. In the group with no incentives 56% of choices of *P* bets were inconsistent with the valuations. In the group facing incentives that proportion was 70%. From this the surprised authors conclude that in the group making decisions for real money "the preference reversal phenomenon is not only replicated, but is even stronger" (p. 632).

The outcomes of the bets in Grether and Plott's experiment ranged from –$2 to $40, and their expected values, from $1.35 to $3.86. Pommerehne et al thought that these amounts were too small to motivate subjects. So they multiplied the nominal amounts by 100, while substituting Swiss francs for US dollars. However these amounts were just "play money." Subjects' actual payments were shares of a total of 2000 real Swiss francs determined on a pro rata basis according to their "play money" earnings. As there were 84 subjects, actual individual earnings averaged SFr 23.8, far more than one would expect to earn in Grether and Plott's experiment. However the authors seem to put their hope more on the zero-sum game nature of the payment scheme, and on monetary illusion, which is ironic given that the aim is to motivate rational individual decisions.

Apart from these different incentives, the experiment basically followed Grether and Plott's design, so that the results could be compared. As in previous experiments, standard reversals were frequent, and non-standard ones, few. But as the authors had expected the frequency of standard reversals was lower than in Grether and Plott's experiment: 45% of choices of *P* bets were reversed in valuations in Pommerehne et al., against 67% in the Grether and Plott's experiment.[4]

In each pair both bets had quite similar expected values: the highest exceeded the lowest at most by 5%. Pommerehne et al thought that this might bore subjects and decrease their motivation. So they ran an experiment to assess the effect of increasing the difference between expected values of the bets in each pair. They conclude that different expected values decreased the frequency of preference reversal but that the decrease is not statistically significant.

Reilly (1982) observed in casual talks with subjects who had participated in an earlier experiment that several of them had not perceived the gains and losses as real, although they had been assured that they were. Therefore in a further experiment he placed subjects' $4 credit on their desks right at the beginning of the experiment, and told them that any gain or loss would be added to or subtracted from that, and that the resulting amount would be paid at the end of the experiment. The usual practice had been to hand subjects any money only at the end of the experiment. This change seems to have reduced the frequency of preference reversals, but the exact effect is

[4] This percentage differs from the one reported earlier because it concerns only four of the six pairs in Grether and Plott's experiment, the ones used by Pommerehne et al.

hard to disentangle from those of other changes introduced at the same occasion in the experimental design.

These experiments tried to test a couple of other possible causes of preference reversal.

Grether and Plott's second experiment aimed at testing the possibility of preference reversals being caused by an irrational bargaining behaviour. In most, if not all, real situations sellers facing a prospective buyer, if asked, will state a higher price than the minimum they are prepared to accept, in the hope of influencing upwards the buyer's offer. When valuations are elicited using the BDM procedure the buyer's offer is random, thus there is nothing subjects can do to influence it. By overstating their true certainty equivalents they only risk forgoing an offer they would be willing to accept. However the question "what is the minimum price you would sell this bet for" might trigger a seller instinct, and overcome these rational considerations. To test this the authors used a variant of the BDM procedure that instead of asking subjects to state their minimum selling price asked them to state "the exact dollar amount such that you are indifferent between the bet and the amount of money." The frequency of preference reversals thus obtained was not much different from those obtained with selling prices. From that the authors conclude that bargaining behaviour cannot explain preference reversals.

Pommerehne et al explored the possibility of preference reversals being the result of inexperience. Their experiment had two runs. In each run subjects performed the full set of valuation and choice tasks, and after the first run they played one of their decisions for real before proceeding to the second run. The authors conclude that the experience obtained in the first run somewhat reduces the frequency of reversals in the second run, but that the reduction is not statistically significant.

Reilly assesses the effect of providing subjects the expected value of the bets. He finds the reduction in the frequency of reversals resulting from that to be statistically significant, although the frequency stays quite high.

4. How to Measure Preference Reversal

In the four studies reviewed so far standard reversals are far more frequent than non-standard reversals. One could take the view that each single inconsistency between valuation ranking and choice is a violation of the economic theory of rational choice, and disregard the breakdown of inconsistencies between standard and non-standard reversals. The view taken on this thesis is that, on the contrary, a degree of randomness may exist in decision making, and thus unsystematic inconsistencies are not incompatible with optimising behaviour. Thus the asymmetry between standard and non-standard reversal is what defines preference reversal and challenges the economic theory of rational choice.

This view has been expressed, for instance, by Cox and Grether (1996) and Plott (1996), but it is not clear that a consensus on the issue exists among economists. For instance Berg et al (1985) focus on the sum of all inconsistencies, and Harrison (1994) does not even report the breakdown of inconsistencies by type. The discoverers of preference reversal, Lichtenstein and Slovic (1971), stress the asymmetric pattern of the phenomenon, and see it as a confirmation of their views. Their compatibility hypothesis predicts standard but not non-standard reversal. Non-standard reversal, the authors say, "might best be thought of as a result of carelessness or changes in *S*[ubject]'s strategy during the experiment" (p. 48); randomness in decision making, one might say.

If randomness in decisions gives rise to non-standard reversal, it should also give rise to standard reversal. Therefore reversals are a problem for the theory of rational choice only to the extent that their pattern cannot be explained by some stochastic element in decision making. At first sight the asymmetry between standard and non-standard reversal observed in the above-reviewed studies seems greater than what randomness could possibly explain. But how much asymmetry, if any, is compatible with randomness alone? And how should that asymmetry be measured? Should we take the difference between the rates of standard and non-standard reversals as a measure of the strength of preference reversal? Even if the answer to the last question is yes, we face a problem: standard and non-standard reversals can be measured by three different rates, and often these rates present different pictures.

A second look at the studies reviewed in the previous section will make clear the importance of the measurement issue. Data analysis in these studies is based on one type of rate only. The data are partitioned according to whether the *P* bet or the *$* bet is chosen. Then standard reversals are measured as a proportion of the number of choices of *P* bets, and non-standard reversals as a proportion of the number of choices of *$* bets. That is what I will call rates of reversal conditional on choices. No explicit justification for this measure has been presented. However there are two other ways, at least, of measuring preference reversals. These often lead to conclusions that are quite different from those of the authors. One could measure preference reversals as a proportion of the number of total choices made by subjects. That is what I will call unconditional rates of reversals. This measure was used by Harrison (1994) for instance. And one could also measure standard reversals as a proportion of cases of the *$* bet being valued more highly than the *P* bet, and non-standard reversals as a proportion of the opposite cases. That is what I will call rates of reversal conditional on valuations.

Grether and Plott's experiment 1 provides an example of how these three measures lead to different conclusions. Table 2 summarises the results of this experiment. Subjects' responses are shown as proportions of the total number of choices. Therefore the unconditional rates of reversal can be read directly from this table.

Table 2: Responses, %, in Grether and Plot: The Effect of Incentives

	No Incentives			With Incentives		
	Highest Price			Highest Price		
Choice	*P*	*$*	Total*	*P*	*$*	Total*
P	20.0	**29.0**	49.0	9.9	**26.4**	36.3
$	**5.7**	45.3	51.0	**8.4**	55.3	63.7
Total*	25.7	74.3	100.0	18.3	81.7	100.0

* Totals exclude cases of indifference and equal prices.
Bold numbers denote preference reversals.
Source: Grether and Plott (1979), tables 5 and 6.

The other two measures, rates conditional on choices and rates conditional on valuations, can be computed from table 2. All three types of rates are shown in table 3. Each of the three types of rate tells a different story, and only the rates of reversals conditional on choices support the authors' assertion that incentives make the preference reversal phenomenon stronger. Unconditional rates show a decrease in the reversal asymmetry. Rates conditional on valuations show the disappearance of the usual asymmetry. Does that allow one to conclude that preference reversal may no longer be there?

Table 3: Reversals in Grether and Plott's (1979): The Effect of Incentives

Type of Reversal	Experiment 1: Reversal rates (%) conditional on*					
	Choices		Valuations		Unconditional	
	N Inc	Inc	N Inc	Inc	N Inc	Inc
Standard	59.2	72.6	39.0	32.2	29.0	26.4
Non-standard	11.2	13.1	22.2	45.8	5.7	8.4

"Inc" stands for With Incentives; "N Inc" for No Incentives.
* See text above for explanation.
Cases of indifference and equal pricing were ignored. That is why the rates of reversals conditional on choices in this table are slightly higher than the ones reported by the authors.
Source: Grether and Plott (1979), tables 5 and 6.

Pommerehne et al (1982) claim to obtain a 22 percentage point reduction in (standard) preference reversals relative to Grether and Plott (1979) with their stronger incentives. The reduction turns out to be more modest if rates of reversals conditional on choices are not used. See the table below.

Table 4: Reversals: Pommerehne et al (PSZ) versus Grether and Plott (GP)

Type of Reversal	Reversal Rates (%) Conditional on*					
	Choices		Valuations		Unconditional	
	PSZ	GP†	PSZ	GP†	PSZ	GP†
Standard	48.1	70.8	34.2	36.1	23.3	29.0
Non-standard	13.1	13.5	21.2	40.0	6.7	8.0

*See text above for explanation.
†These rates differ from those in table 3 because they concern only four pairs of bets, the ones used by Pommerehne et al.
Cases of indifference and equal pricing were ignored. That is why the rates of reversals conditional on choices in this table are slightly higher than the ones reported by the authors.
Source: Pommerehne et al (1982), table 2.

Pommerehne et al (1982) provide yet another example of diverging measures. Table 5 compares group I subjects' responses in both runs of their experiment. The authors conclude that the experience subjects may have acquired in the first run reduced the reversal rate, but not significantly. This conclusion is based on the rates of reversal conditional on choices. But if we look at the unconditional rates we see that many subjects who reversed in the first run did not in the second.

Table 5: Responses (%) in Pommerehne et al: effects of experience in group I

	Reversal Rates (%) Conditional on*					
Type of	Choices		Valuations		Unconditional	
Reversal	1st run	2nd run	1st run	2nd run	1st run	2nd run
Standard	54.2	47.5	48.6	25.5	32.5	18.2
Non-standard	14.0	13.6	16.9	29.5	5.6	8.4

*See text above for explanation.
Cases of indifference and equal pricing were ignored. That is why the rates of reversals conditional on choices in this table are slightly higher than the ones reported by the authors.
Source: Pommerehne et al (1982), table 3.

If the distribution of choices and highest valuations between *P* bets and *$* bets remained constant when the experiment design changes the relative changes in all the three types of rates would be the same. It is because these distributions change that the three types of rates change differently. For instance in Grether and Plott's (1979) experiment 1, incentives caused choices of *P* bets to decrease from 49% to 36.3% of total choices. Therefore standard reversals as a proportion of choices of *P* bets increased, from 59.2% to 72.6%, although as a proportion of total choices they decreased slightly, from 29% to 26.4%. One might ask whether it is legitimate to say in this case that the preference reversal phenomenon became stronger.

There might be a case for using rates of reversals conditional on choices instead of any of the other two if choices reflect preferences, and the cause of preference reversals lies mainly in biased pricing. However if the cause of preference reversals lies mainly in choices the rates of reversals conditional on valuations might be a more appropriate measure. In the absence of generally accepted, explicit hypotheses about the cause of preference reversals it may be difficult to tell whether preference reversal becomes stronger or weaker if the different types of rate move differently.

If one accepts that randomness exists in decisions the strength of preference reversal and even its very existence would best be evaluated by reference to a stochastic model of rational choice and valuation. Lichtenstein and Slovic (1971) tested and rejected a model of errors in valuation ranking and choices. Chapter 4 shows that the authors actually tested a particular set of parameters, not the general model. To my knowledge this is the only attempt to fit a stochastic model of rational choice and valuation to preference reversal data. A number of stochastic models of rational choice, Harless and Camerer (1994), Hey and Orme (1994), and Loomes and

Sugden (1995), to be reviewed in chapter 4, have been proposed, but these are models of choice only, not valuations, and are not directly applicable to preference reversal data. These models show however that random elements in decision making can give rise to deviations from the predictions of rational choice models that appear non-random at first sight. Thus the extent to which preference reversal deviates systematically from the theory of rational choice remains an open question.

Chapter 4 is an attempt to help answer that question. It revisits Lichtenstein and Slovic's (1971) model, and develops a random preference model of rational choice and valuation. This new model is based on Loomes and Sugden's (1995) theory of random preferences. The two models are then tested against several datasets.

5. Does it Pay not to Reverse?

The incentives question was picked up again by Harrison (1994). He notes that to elicit subjects' preferences an experimental design must meet a *dominance condition*: the rewards from making accurate decisions must dominate, that is, be higher than the cost arising from the mental effort required to make those accurate decisions. Then he argues that in preference reversal experiments the gains subjects forwent by making inconsistent decisions were so small that subjects had hardly any incentives to incur the additional subjective costs required for better decisions.

He analysed the data from Reilly's (1982) experiment, and found that the opportunity cost per decision due to inconsistent decisions was on average only 0.64 cents. He does not report the inconsistency costs in Grether and Plott's (1979) experiment, but in his replication of it, in which he found qualitatively identical results, the cost per decision was 0.6 cents. Costs per decision take into account that each decision has only a 1 in 18 chance of being played for real. If one thinks that the relevant cost of an inconsistent decision is the cost if that decision is played for real the costs would be about 12 cents. For comparison, the expected value of the bets ranged from $1.67 to $3.85.

Taken at face value, Harrison's critique calls into question the relevance of observed reversals. Yet his critique seems to have motivated little reaction from experimentalists. There is however at least one reason why Harrison's critique should not be taken at face value: the opportunity costs of inconsistencies are not observable.

The concept of opportunity cost of the misreport of an expected utility maximiser's preferences is easy to grasp, and Harrison (1992) explains how to compute them theoretically. Suppose the BDM procedure is used to elicit certainty equivalents, and a subject reports a valuation for a bet in excess of his true certainty equivalent. If the random offer price is lower than the certainty equivalent or higher than the reported value, the outcome to the subject is the same as if the true certainty equivalent had been reported, so the misreport brings no cost. If the offer price, x, happens to lie between the certainty equivalent, ce, and the valuation, v, the subject will keep the lottery instead of receiving the preferred offer price, so the opportunity cost is $x - ce$. Denoting $p(x)$ the probability function used to generate the offer price, and measuring all variables in cents[5], the expected cost, *ECost*, from a misreported certainty equivalent is

$$ECost = \sum_{x=ce}^{v-1} (x - ce)p(x) \qquad \text{for } v > ce. \tag{3}$$

Obviously the expected costs of misreports cannot be computed because certainty equivalents are not observable. Harrison (1994) assumes the valuations of *P* bets and choices to reflect the true preferences, so that whenever there is a reversal the certainty equivalent of the *$* bet must have been misreported. Under these assumptions a standard reversal implies that the certainty equivalent of the *$* bet, $v_\$$, is lower than the valuation of the *P* bet, v_P. Therefore substituting v_P for $ce_\$$ in the expression above we obtain the minimum of the possible expected opportunity costs, *MEC*, of misreporting the certainty equivalent of the *$* bet:

$$MEC = \sum_{x=v_P}^{v_\$ -1} (x - v_P)p(x) \qquad \text{for } v_\$ > v_P. \tag{4}$$

MEC depends only on observable variables, therefore I suspect this is what the author actually computed. This, if true, weakens the author's argument, as *MEC* may grossly underestimate the true opportunity cost, as the figure below illustrates.

[5] Or in the smallest monetary unit used in the offer prices, usually one hundredth of the standard monetary unit.

As the random offer price is uniformly distributed, the above expressions are proportional to the shaded areas: the darker triangle represents only the visible cost, the computable *MEC*; but the cost that is hidden from the experimenter, the lighter coloured trapezoid, can be larger than *MEC*. For example, if the difference between v_P and $ce_\$$ is the same as the difference between $v_\$$ and v_P, *ECost* is four times as big as *MEC*.

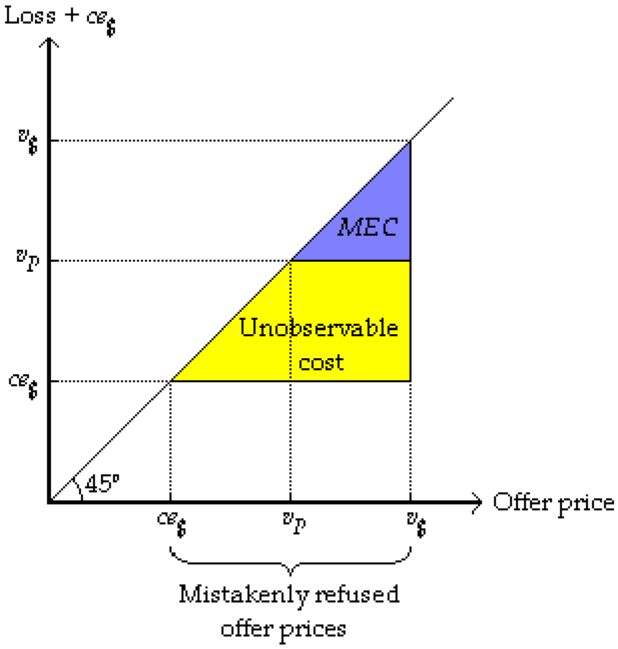

Figure 1: Expected opportunity cost (*Ecost*) of the misreport of a *$* bet is the sum of *MEC* and an unobservable cost

The author not only underestimated the opportunity cost of a preference reversal, as also assumed that subjects incurred no cost when their valuations were consistent with their choices. This assumption may often be violated, as subjects incur an expected opportunity cost whenever their valuations differ from their certainty equivalents, no matter whether their valuations are consistent with their choices or not. Therefore the opportunity costs of inaccurate decisions may have been substantially higher than the figures reported by Harrison.

Harrison (1994) claims to have shown that the frequency of preference reversals greatly decreased when he changed the experiment design to meet the dominance condition.

In one of his experiments he changed the bets of Grether and Plott's experiment so that the expected value of the *P* bet was higher than that of the paired *$* bet. The ideas was that even if subjects commit misreports of, say, up to 50 cents (because the opportunity cost of doing that is small) and if certainty equivalents differ by one dollar, then the ranking implied by pricing will always express the true preferences. When the difference between expected values was 100 cents reversals of both types as a proportion of total choices were only 10%; they were 15% when the difference was 50 cents; in the control group, using the same bets as Grether and Plott, with pairs made of bets with similar expected values, they were 45%. This reduction in the frequency of reversals is hardly surprising: if the *P* bets were made sufficiently more attractive than the paired *$* bets, subjects would always choose the *P* bet and value it more highly as well. As for the author's remark that (p. 239) "a disparity of only 100 cents was sufficient to induce consistent behaviour", it should be noticed that these 100 cents was between 26% and 79% of the expected value of the *$* bets.

In another experiment subjects had to report valuations in multiples of 25, 50, 100 or 200 cents. The idea was to increase the opportunity cost of misreports. When multiples of 100, and 200 cents were used reversal rates were about 30% and 20% respectively, against 45% in the control group. This is not very convincing. As Camerer (1995 p. 662) says, "if subjects could report only one price then no reversals would occur."

In another experiment Harrison reduced the range of the offer price to between zero and 499 cents, keeping all other features of the control group. This doubles the expected opportunity cost of any given misreported valuation. The rate of preference reversals was basically the same as in the control group. So were the opportunity costs of inconsistencies reported by the author. This means that the magnitude of preference reversals must have been smaller.

It would be interesting to have used in a control group an offer price range from -500 to 499 cents, to tell the effect of the increased opportunity costs from the effect of simply decreasing the largest valuation subjects can report. It would be hardly surprising if most people failed to notice the effect of the offer price range on the opportunity cost of misreporting valuations. If subjects perceived the opportunity costs as anything bearing any resemblance to expression (3) above they could as well perceive what the right decision was, as Harrison (1992) points out.

This means that if low opportunity costs cause lack of dominance the reason is probably not that subjects perceive costs to be small, but that they do not perceive costs at all, and thus do not notice they are misreporting. The fact that in most experiments subjects play just one of their decisions at the end of the experiment does not help make costs of possible mistakes apparent. The way to increase cost awareness is probably by playing out all decisions with immediate feedback during the course of the experiment,[6] rather than changes in design aiming at increasing the expected cost of any misreport.

6. Beyond Gambles: Compatibility and Prominence

Preference reversal in gambling decisions can be seen as an instance of a broader discrepancy between choice and matching. This discrepancy involves pairs of options described by two attributes. Subjects are more likely to express preference for the option that is superior on the more important of the attributes in choices than in the logically equivalent matching tasks. The occurrence of this phenomenon in a variety of contexts was first demonstrated by Slovic (1975), and has been confirmed by, for instance, Tversky, Sattah, and Slovic (1988), Fischer and Hawkins (1993), and Fischer, Carmon, Ariely, and Zauberman (1999). Consider the following example, which is part of a study conducted in Israel by Tversky et al (1988, p. 373).

> About 600 people are killed each year in Israel in traffic accidents. The ministry of transportation investigates various programs to reduce the number of casualties. Consider the following programs described in terms of yearly costs (in millions of dollars) and the number of casualties per year that is expected following the implementation of the program.

	Expected number of casualties	Cost
Program *X*	500	$55M
Program *Y*	570	$12M

A group of subjects was asked to choose their preferred program. Another two groups of subjects were presented the same problem, but with one of the costs missing. They were asked to state the missing cost that would make the two programs equally attractive. This is a matching task. Suppose that the cost of program *X* is missing, and a subject indicates that a program that reduces the number of deaths to 500 at a cost of, say, $60M is as attractive as program *Y*. Assuming

[6] This is compatible with the random lottery selection: subjects could be informed at the beginning of the experiment that only one decision would count for real.

monotonicity, she should prefer to reduce the number of deaths to 500 at a cost of \$55M, program *X*, than at a cost of \$60M. Thus, assuming transitivity, she should prefer *X* to *Y*. Generally a subject that indicates a cost above \$55M for a program that reduces the number of deaths to 500 should prefer *X* over *Y*; a subject who indicates a cost below \$55M should prefer *Y* over *X*. Preferences can be inferred in the same manner when the missing cost is that of program *Y*. The preferences inferred from the matching costs indicated that only 4% of subjects preferred program *X*. In the choice group this program was preferred by 67% of subjects.

This discrepancy between choice and matching is akin to preference reversal. A monetary valuation of a gamble in preference reversal experiments is a matching task. It can be seen as the matching of two gambles. One is the gamble being valued, and the other is a yardstick gamble, that offers an unspecified payoff with certainty. The subject is asked to state the unspecified payoff that would make both gambles equally attractive to her. Thus preference reversal is also a discrepancy between choice and matching. Both discrepancies imply violation of either transitivity or procedure invariance. At first sight there seems to be only one difference between the two phenomena. In preference reversal experiments it takes two matching tasks to infer the preferences on the paired options; in the choice-matching discrepancy it takes only one. This difference may be relevant if one seeks to explain the discrepancies by random errors occurring independently in each task, but otherwise it does not seem important.

Similar, although less extreme, choice-matching discrepancies were observed in a variety of contexts. Tversky el al (1988) observed the discrepancy in the following problems: two job applicants, one with better technical knowledge but less socially skilled than the other; another pair of job applicants, one more competent but less creative than the other; two beach clean-up programs, one more comprehensive but also more costly than the other; two reward packages, one offering more money in the short term but less in a later year than the other; and another two reward packages, one offering more book coupons but fewer travel coupons than the other. Slovic (1975) found the discrepancy in such problems as typists, one faster but less accurate than the other; car tyres, one brand better but more expensive than the other; or television adverts, one more frequent but shorter than the other (to be evaluated according to annoyance). Other objects that induced the discrepancy include flats (Fisher and Hawkins 1993), jobs, binoculars, and lawyer services (Fischer et al 1999).

The discrepancy is not random. In all problems the paired options were described by two attributes. The option that scores better in the straight choice than in matching is always the option that is judged superior on the prominent attribute, that is, the more important of the two attributes. This has been called the *prominence effect*. In some problems the authors thought that one attribute would emerge naturally as the prominent one. For instance in the road-safety example, the authors assumed that most subjects would deem the number of lives saved more important than the cost of the program. If that assumption is right the prominence effect emerges: program *X* saves more lives, and was preferred by more people in the choice task than in the matching tasks. In some problems there is no compelling reason to see one attribute as naturally more important than the other. In those cases the description of the problem would explicitly state which attribute was more important. For instance, in one problem involving job applicants, some subjects were told that competence was more important than creativity, and others were told the opposite. More subjects preferred the candidate superior on the prominent attribute (regardless of whether it was competence or creativity) in the choice task (65%) than in the matching task (38%).

Note that preference reversal conforms with the prominence effect if the probability of the best outcome is the prominent attribute in gambles. In that case there would be a greater tendency to prefer the *P* bet in the choice task than in the valuation tasks.

In the problems involving road-safety or beach clean-up programs, the missing value was always the cost of one of the programs, never the value of the other attribute. In the other problems each of the four values was missing for some group of subjects. The choice-matching discrepancy was observed regardless of which value was missing. Thus, while preference reversal could be seen as a tendency, misleadingly strategic or otherwise, to overstate the values of the lotteries (there is ample room for overstating the value of *$* bets, but not that of *P* bets), the choice-matching discrepancy could not result from such a tendency. For instance in the road-safety example, overstating the cost of program *X* implies preference for *X*, whereas overstating the cost of *Y* implies preference for *Y*. A general tendency to understate a value will also imply different preferences depending on which value is missing.

The prominence effect is founded on the relative importance of two attributes. But what does it mean to say that one attribute is more important than the other? The answer may be easy and precise if both attributes are measured on the same scale. For instance if technical knowledge and social skill are both rated on a scale from 40 (very poor) to 100 (superb), as in Tversky et al (1988), one may view technical knowledge as more important than social skill if one is willing to forgo more than one point in social skill to obtain an additional point in technical knowledge. But what does it mean to say, for instance, that life is more important than money? Surely I would part with all my material wealth rather than lose my life, but this is of little help in defining the relative importance of money and life in the evaluation of road-safety programs. For instance, what cost per life saved implies giving equal importance to life and money?

Nevertheless, despite the fuzziness of the concept of relative importance in some contexts, psychologist have been able to use it to predict which option fares better in choice than in matching.

Tversky et al (1988) see the prominence effect as a violation of procedure invariance, namely as resulting from the use of different heuristics in choices and matching tasks.

They suggest that subjects may resort to the following heuristics to make their choices. Suppose a person is given a choice between two options, each described by the same two attributes. She could start by checking whether one option dominates the other, that is, whether one option is superior on one attribute and no worse than the other option on the other attribute. If that were the case, the choice would be obvious. If not, she could check whether any option has such a large advantage in one of the attributes that it obviously more than compensates the disadvantage in the other attribute. If so, the choice would also be easy. If not, an easy way of making the choice would be to choose the option that is superior on the prominent attribute. There are alternative strategies. The person could work out the rate at which she is willing to trade one attribute for the other, for instance, how much to pay to save an extra life, and compare it with the rate implicit in trading one option for the other. This exercise requires effort, and it is difficult to find a compelling argument to justify a particular subjective rate of substitution. On the contrary, if one of the attributes stands out as more important than the other, relying on it is easy and provides a compelling justification for one's choice.

Tversky el al (1988) argue that this strategy is not feasible in matching tasks. This strategy relies on ordinal comparisons of attribute values across options and of importance of attributes. That makes it appropriate to arrive at an ordinal comparison of two options, but not to arrive at a precise value, which is what one must do in a matching task. To do this one must establish one rate of exchange between attributes, or find what difference across options between the values of one attribute matches the difference between the values of the other attribute. Tversky et al (1988) argue that in doing this people tend to undervalue the difference in importance between the two attributes.

The consequence of these two strategies is that the prominent attribute is given more weight in choice than in matching. Tversky et al (1988) see this prominence effect as resulting from the general principle of compatibility. This principle states that the weight of an input is larger the more compatible it is with the output. The output in a choice is a differentiation between two options (a point made explicit by Fischer et al 1999), which is compatible with differences in attribute value across options, and with differences in importance of attributes. This compatibility highlights the difference in importance of attributes, increasing the relative weight of the more important attribute. In contrast, matching, which does not differentiate between options, does not highlight the difference in importance of attributes. This was later termed the *strategy compatibility hypothesis* by Fischer and Hawkins (1993), and has been thus summarised: the choice, an ordinal task, is compatible with ordinal strategies; matching, a cardinal strategy, is compatible with cardinal strategies.

We have encountered the compatibility principle before, in section 2 of this overview of preference reversal. Lichtenstein and Slovic (1971) suggested that because the amount to win in a gamble and the monetary valuation are both expressed in dollars and because there is no obvious difference in compatibility between the choice and either amount to win or probabilities, the amount to win is given more weight in valuations than in choices. This was later termed the *scale compatibility hypothesis* by Fischer and Hawkins (1993), to distinguish it from strategy compatibility.

In some circumstances the scale-compatibility hypothesis predicts the same results as the strategy-compatibility hypothesis; in other circumstances it does not.

The output of a matching task is necessarily expressed in the same unit as one of the attributes. Therefore there is always scale compatibility.

If the two attributes are measured on different scales, the attribute to which the missing value refers is compatible with the output of the matching task; the other attribute is not. If the missing value refers to the prominent attribute, scale compatibility should enhance its weight in the matching task relative to that in the choice task. This is the opposite of what the strategy-compatibility hypothesis predicts. If the missing value refers to the secondary attribute the weight of the prominent attribute, according to the scale-compatibility hypothesis, is reduced in the matching task relative to that in the choice task. This is also what the strategy-compatibility hypothesis predicts.

For instance, in Tversky et al's (1988) road-safety problem, the missing value referred always to the cost, the secondary attribute. Therefore the output of every matching task was compatible with the cost attribute. Then, according to the scale-compatibility hypothesis, the cost should receive more weight in matching than in choice, which accounts for the observed behaviour in that problem. If the missing value referred to the number of lives saved scale compatibility alone would increase the weight of lives saved in matching relative to choices. Thus scale compatibility could not give rise to the prominence effect, whereas strategy compatibility would.

In some problems both attributes are measured in the same scale. For instance in Tversky et al's (1988) problems involving job applicants, technical knowledge and social skill in one case, competence and creativity in the other, are all rated on a scale from 40 to 100. Thus there is scale compatibility between both attributes and the output of every matching task regardless of whether the missing value refers to the prominent or to the secondary attribute. This could blunt the impact of differences between the importance of attributes in the matching task, and thus produce the prominence effect. The case here is however less clear cut than when the scales are different and the missing value refers to the secondary attribute.

It may be useful to distinguish what one may call *content compatibility* from scale compatibility. For instance one would expect technical knowledge to be more compatible with technical knowledge than with social skill, even if both are rated on the same scale. Thus, according to the general compatibility principle, technical knowledge should receive a bigger weight when the missing value refers to itself than when the missing value refers to social skill. Tversky et al (1988) report that in the problems involving job applicants preferences inferred from matching show a higher proportion of subjects preferring the option that is superior on the prominent attribute when the missing value refers to this attribute than when it refers to the secondary attribute. This could not result from scale compatibility, but could result from content compatibility.

All of Tversky et al's (1988) results could possibly result from scale compatibility: either the attributes were measured on the same scale or the missing value in the matching tasks referred always to the secondary attribute. That was not the case in Fischer and Hawkins (1993). They observed the prominence effect in problems involving prizes, one offering more money but fewer days of paid holidays than the other; and flats, one more expensive but closer to campus than the other. The prominence effect was observed when the missing value referred to the attribute they considered prominent, money prize or rent. This contradicts the scale-compatibility hypothesis, but is consistent with the strategy-compatibility hypothesis. The rejection of the scale-compatibility hypothesis in the prize experiments is especially interesting, as preferences were inferred from choices and a pair of monetary equivalents. Thus the prize experiments followed the design typical of a preference reversal experiment, and scale compatibility has been offered as an explanation for preference reversal.

In one of these experiments Fischer and Hawkins (1993), in addition to the choice and pricing tasks, also asked subjects to indicate their strength of preference in a scale from –5 to 5. The experiment had a within-subjects design, and subjects performed the strength-of-preference and the choice tasks in sessions at least one day apart. With the strength-of-preference task the authors wanted to test a simple interpretation of the strategy-compatibility hypothesis, namely whether merely asking subjects to specify a value would induce them to follow cardinal strategies, and give less weight to the prominent attribute than in choices. That was not the case: subjects were as likely to prefer the prize offering the higher amount of money in this task as in the choice task. A possible explanation is that, having the two prizes fully described, subjects decided first which one they preferred, and decided only afterwards how much they preferred it to the other.

Fischer et al (1999) pointed out that in a choice one differentiates two options, whereas in a matching task one equates two options. This distinction fits neatly into Tversky et al's (1988) suggested explanation for the prominence effect, namely that choices highlight differences between importance of attributes. Fischer et al (1999) conjectured that the prominence effect could also be observed in other differentiating and equating tasks. This reasoning led to their *task-goal hypothesis*, which proposes that the prominent attribute is given more weight in tasks where the perceived goal is differentiation than in tasks where the perceived goal is equation.

Two of their studies involved jobs characterised by annual salary and duration of holidays. Salary was assumed to be the prominent attribute. Preferences were elicited in choice, matching, and choice-based matching tasks. In the choice-based matching task subjects were confronted with a sequence of choices between two jobs, say, job A and job B. The attribute values of job A did not change. The salary of job B kept changing so as to converge towards the value that made the subject indifferent between the two jobs. The sequence of choices would stop when that indifference-inducing salary was within a known interval of USD 100. The aim of the choice sequence was clear to subjects, especially as they matched eight pairs of jobs. In one study the proportion of preferences for the high-salary job inferred from choice-based matching (37%) fell between the proportions inferred from matching (23%) and from choice (63%). The authors conclude that the individual choices in the choice-based matching sequence were influenced both by the differentiating goal in each choice and by the perceived equating goal of the whole sequence.

In a second study preferences were elicited with two versions of choice-based matching. In one the instructions made the end goal even clearer than in the first study. In another, that goal was hidden. The eight choice sequences were interwoven together. As some sequences converged faster than others, filler choice tasks were introduced if necessary so that at least five tasks would mediate two consecutive choices of any individual sequence. The proportion of preferences for the high-salary job inferred from the hidden choice-based matching (52%) was even slightly higher than that inferred from choice (48%), whereas the proportion inferred from the transparent choice-based matching (34%) was about the same as that inferred from matching (36%). It seems as though subjects, expecting a long series of choices in the transparent choice-based matching, decided on a matching salary, and used it to answer expediently all the choices in a sequence.

In these two studies the missing value in the matching task referred to the prominent attribute, and the two attributes were not measured in the same scale. Thus these studies offer further evidence contrary to the scale-compatibility as an explanation for the prominence effect.

A third study consisted of problems involving airport shuttles, AM/FM cassette players, binoculars, lawyer services, and the problems involving road-safety and beach clean-up programs we saw in Tversky et al (1988). The attributes were cost and quality. Quality was assumed to be the prominent attribute. In addition to the choice and matching tasks, subjects answered three versions of high-low questions. Consider for instance the road-safety problem and a matching task where the cost of program *X*, $55M, is missing. Subjects would be asked whether a cost of $55M is too high or two low to make both programs equally attractive. This is the basic high-low question. In another version, high-low, match now, subjects were asked the previous question and the standard matching question. Both questions were displayed simultaneously. In a third version, the high-low, match later, subjects were presented the high low question and had been informed that later they would be asked the matching question. The missing value referred always to the secondary attribute. The results reported are based on the answer to the high-low question, not to the following matching question. The proportion of preferences for the option superior on the prominent attribute were similar when inferred from choices (51%) and the basic high-low task (53%); it was lower when inferred from the high-low, match later task (43%), even lower when inferred from the high-low, match now task (33%), and lowest when inferred from matching (15%).

The basic high-low question appears to have been answered as a straight choice. In the other two versions of the task, the presence of the matching question influenced the answer to the high-low question.

This literature clearly shows that formally different but logically equivalent preference-eliciting procedures give rise to systematically different revealed preferences. The response pattern is akin to preference reversal. The explanation psychologists originally suggested for preference reversal, scale compatibility has been rejected in several studies. It may be the case that the explanation for preference reversal is strategy compatibility not scale compatibility. If probabilities are the prominent attribute in gambles, the scale and strategy-compatibility hypotheses predict the same results in the typical preference reversal experiment, where monetary valuations were used.

The literature reviewed in this section must be seen with some reservation though. Subject's answers had no real consequences. Moreover, some questions do not clearly ask for a specific type of decision, even if hypothetical. For instance in

Fischer et al (1999) the choice task of study 3 asks "Assuming you must choose one of the two, which one would you select?" When the choice is between products or services characterised by price and quality, it is not clear whether the subject should choose the option he would buy or the option he would prefer to be given. The former interpretation makes more sense, otherwise the choice would obviously be the high-quality option, but nothing clearly precludes the latter interpretation. Tversky et al (1988) in the matching task of the job applicants problem ask subjects "to complete the missing score so that the two candidates are equally suitable for the job." This instruction is clear, as it specifies one criterion, suitability for the job. The equivalent question in the road-safety problem (these are the only questions reported in the paper) is not as clear: "you are asked to determine the cost of Program *X* that would make it equivalent to Program *Y*." It is not clear what the criterion for equivalence should be. It is almost too tempting to think of equivalence in terms of cost per life saved.

Some results in Tversky et al (1988) seem to show signs of confusion or carelessness. In some problems subjects answered questions in a high-low, match-now task, as in Fischer et al (1999). A fair proportion of subjects gave contradictory answers in the two questions of the task, for instance indicating that a cost of $55M was too low to make the two options equivalent, and then indicating less than $55M as the appropriate cost. Note that the two questions were displayed together and answered at the same time. Most of the inconsistencies, 86%, followed the prominence effect pattern. The unweighted average across four problems of the proportion of preferences for the option superior on the prominent attribute was 48% when inferred from the high-low answers, and 30% when inferred from the following matching answer. As some subjects made the opposite type of inconsistency, around 20% of all subjects gave contradictory answers in the same task (the exact proportion cannot be computed from the published data, as some subjects appear to have answered one part of the question but not the other).

The prominence effect, as preference reversal, seems to indicate violation of either transitivity or procedure invariance. The strategy-compatibility hypothesis, or its generalisation, the task-goal hypothesis, offers a compelling account of how different response modes may elicit different mental strategies, and lead to different answers, thus violating procedure invariance. However the lack of incentives and the imprecision of some questions may have lowered the impact of subjects' preferences on their answers.

7. Rescuing Procedure Invariance

During the eighties there were several attempts to show that what looks like preference reversal could be the result of optimising behaviour.

The first was made by Loomes and Sugden (1983), with their regret theory. The following table helps to illustrate regret theory as applied to our initial pair of lotteries.

Table 6: Acts' Assignment of Consequences to Events

	Number drawn from a bingo cage		
Choice	1 - 7	8 -35	36
P bet	\$4	\$4	−\$1
\$ bet	\$16	−\$1.5	−\$1.5
\$3.8	\$3.8	\$3.8	\$3.8

Suppose an individual faces a choice between the *P* bet and the *\$* bet. After his choice a ball will be drawn from a bingo cage containing 36 balls numbered 1 to 36.[7] Table 6 shows the individual's gain or loss, the consequences, resulting from his choice and the number drawn, the event. Conventional utility theory assumes that if the individual chooses the *P* bet and wins \$4, his satisfaction, or utility, will be the same regardless of the number drawn from the bingo cage. Regret theory assumes instead that his utility will be higher if the number is from 8 to 35 than if it is from 1 to 7. In the latter case the individual will regret not having chosen the *\$* bet,[8] and that feeling will decrease the satisfaction of having gained \$4; if the number is 8 to 35 the individual will rejoice for having won \$4 rather than lost \$1.5, and that will increase his satisfaction. The authors formalised this idea in what they named *modified utility* function: the modified utility of choosing bet *x* when bet *y* could have been chosen, and event *i* occurs is

$$M(x_i,y_i) = C(x_i) + R[C(x_i) - C(y_i)], \tag{5}$$

[7] As in Grether and Plott's experiments.

[8] I assume the individual sees the number drawn from the bingo cage as independent of the bet he chooses.

where $C(x_i)$ represents the utility if the individual had obtained x_i in a way completely unrelated to any choice he made or might have made, and $R[.]$ represents the regret or rejoicing from having chosen one bet rather than the other. Then the individual will choose the bet that gives him the highest expected modified utility.

The interest of this theory, besides the intuitively appealing idea that the expectation of regret or rejoice may influence decisions, is that it predicts, or at least is consistent with several experimental results that violate expected utility theory, namely preference reversal. For instance suppose that our individual has the modified utility function

$$M(x_i, y_i) = \begin{cases} x_i^{0.9} + 0.5\left(x_i^{0.9} - y_i^{0.9}\right)^{\lambda} & \text{if } x_i \geq y_i \\ x_i^{0.9} - 0.5\left(y_i^{0.9} - x_i^{0.9}\right)^{\lambda} & \text{if } x_i < y_i \end{cases}$$

If λ=1.5 and he is asked to make choices between pairs of options of the table 6, he would exhibit an intransitive choice cycle of the kind observed in standard reversals: when facing the choice between the *P* bet and the *$* bet, he would chose the *P* bet over the *$* bet; would choose the *$* bet over $3.8; and would choose $3.8 over the *P* bet. If he had participated in Grether and Plott's experiment he would have stated a minimum selling price of $3.84, for the *$* bet, and of $3.69 for the *P* bet. If λ=1.2 he would have chosen the *P* bet and stated v_P=$3.79 and $v_\$$=$3.34; and if λ=2 he would have chosen the *$* bet, and stated v_P=$3.48 and $v_\$$=$4.51. Thus regret theory is consistent with several patterns of observed behaviour.

Not all patterns though. Regret theory allows the valuation of *$* above *P* when *P* is chosen, but imposes that in these cases *$* be valued less than the winning amount of *P*. A look at table 6 above will reveal that if someone prefers the *P* bet to the *$* bet then she will also prefer $4 with certainty to the *$* bet: if the number 36 is drawn the rejoicing for not having chosen the *$* bet will be higher in the case of $4 with certainty; for any other number the rejoicing or regret will be the same. As many standard reversals involve valuations of the *$* bet well in excess of the winning amount of the *P* bet, regret theory cannot be the sole explanation of preference reversal.

Charles Holt (1986) showed that in experiments using the random lottery selection method and the BDM procedure optimising behaviour may generate apparent preference reversals if the independence axiom of preferences over lotteries does not hold.

Here is a simplified version of Holt's explanation. He restricted his analysis to the case in which subjects face only one pair of bets. Then subjects choose either the *P* or the *$* bet, and set minimum selling prices for both bets. When they set a minimum selling price for a bet they define another lottery. For example by choosing a minimum selling price for a *P* bet of $4, and if the offer price ranges from zero to 999 cents as usual, the subject is actually defining the compound lottery

$$B(4; P) = (P, \tfrac{400}{1000}; \$4, \tfrac{1}{1000}; \$4.01, \tfrac{1}{1000}; \ldots; \$9.99, \tfrac{1}{1000}),$$

or generally $B(v_P; P)$, where v_P is any selling price set by the subject. The corresponding compound lottery resulting from the valuation of the *$* bet is $B(v_\$; \$)$.

Suppose a subject chose the *P* bet and minimum selling prices of $4 for the *P* bet, and $5 for the *$* bet, thus exhibiting an apparent preference reversal. As she knows that only one of the three tasks she has performed is going to be chosen with equal probability to be played for real she may, indeed should, treat *P*, $B(4; P)$, and $B(5; \$)$ as outcomes of the following compound lottery:

$$[P, \tfrac{1}{3}; B(4;P), \tfrac{1}{3}; B(5;\$), \tfrac{1}{3}]. \qquad (6)$$

Thus assuming that the subject maximises her expected utility, and observing the subject's choice of *P*, v_P=4, and $v_\$$=5 and we may conclude she prefers (6) to

$$[\$, \tfrac{1}{3}; B(4; P), \tfrac{1}{3}; B(5; \$), \tfrac{1}{3}]. \qquad (7)$$

If preferences obey independence we may also conclude that she prefers the *P* bet to the *$* bet, since the choice between (6) and (7) will be independent of their common parts. But if the preferences do not obey independence we may no longer derive that conclusion.

Holt notes that apparent preference reversals may also arise because the minimum selling prices subjects set for a bet in these experiments may be influenced by the pricing of the other bet, and by the choice task. Returning to our example, we may conclude that the subject weakly prefers lottery (6) to any lottery

$$[P, \tfrac{1}{3}; B(4;P), \tfrac{1}{3}; B(v_\$;\$), \tfrac{1}{3}] \quad \text{for all } v_\$.$$

But again if preferences are non independent there might be a price $v'_\$$ (or several prices) other than $v_\$$=5 such that the subject strictly prefers $B(v'_\$; \$)$ to $B(5; \$)$.

Thus if preferences do not satisfy independence, and the random lottery selection method leads subjects to view the three tasks performed on a pair of bets as a single compound lottery, subjects' decisions may differ from the ones they would make if they viewed tasks as separated. In particular the latter could be consistent with a single preference ordering of the original *P* and *$* bets even if the former, the ones experimenters may have observed, on surface were not.

Holt's (1986) contribution has implications beyond preference reversal. It shows that the random lottery selection procedure may fail to elicit the preferences the experimenters wanted to investigate. That is, the preferences on the objects dealt with in each task, as opposed to preferences on lotteries over those objects.

Karni and Safra (1987) showed that the BDM procedure may fail to elicit the true certainty equivalents even if the random lottery selection method is not employed, or, more generally, even if subjects do not view the several tasks as a single compound lottery.

They show this with an example, which uses the pair of lotteries of our initial example,

$$P = (\text{-}\$1, \tfrac{1}{36}; \$4, \tfrac{35}{36}) \text{ and } \$ = (\text{-}\$1.5, \tfrac{25}{36}; \$16, \tfrac{11}{36}),$$

and in which preferences are represented by a particular expected utility function with rank dependent probabilities which obeys transitivity but not independence.

With that utility function the certainty equivalents of lotteries *P* and *$* would be $3.065 and $3.038, therefore subjects would choose the *P* bet. But the expected utility of lotteries $B(v_P;P)$ and $B(v_\$;\$)$ would be maximised with minimum selling prices v_p = $3.43 and $v_\$$ = $4.33 respectively. Therefore a phenomenon the authors name *announced price reversal* would be produced, that would look like a standard preference reversal.

Curiously if the authors had assumed that subjects regarded the choice and elicitation tasks as parts of a single compound lottery, as (6) or (7), an announced price reversal would be produced as well, but it would look like an non-standard preference reversal: subjects would choose the *$* bet, and would set minimum selling prices v_P = $4.00 and $v_\$$ = $3.24.

The possibility that the violation of the independence axiom might cause the BDM procedure to fail to elicit true certainty equivalents is counter intuitive. The reason this failure might occur can be made clear with the following example. Suppose the certainty equivalent of lottery *X* is £2.5, and minimum selling prices and random offer prices can be any whole number between £1 and £5. Then if preferences satisfy independence the lottery

$(X, \frac{2}{5}; £3, \frac{1}{5}; £4, \frac{1}{5}; £5, \frac{1}{5})$

will be preferred to lottery

$(X, \frac{2}{5}; X, \frac{1}{5}; £4, \frac{1}{5}; £5, \frac{1}{5})$,

since they only differ in that the second lottery offers an extra 20% probability of obtaining *X* where the first offers £3, and £3 is preferred to *X*. But if the independence axiom does not hold it might be possible that the second lottery be preferred to the first even if £3 is preferred to *X*. In this case the minimum selling price would not be £3, which it would be if preferences obeyed independence.

Indeed Karni and Safra show that if the independence axiom does not hold the BDM procedure will always fail to elicit the true certainty equivalents of some lotteries.

Segal (1988) showed that the preference reversal phenomenon might be caused by a failure of the compound lotteries axiom. He assumes subjects to take the choice task and each of the valuations as independent of each other, as in Karni and Safra (1987). Thus preference reversal arises again from a failure of the BDM elicitation method. He presented an example of a hypothetical subject maximising a utility function that obeys transitivity and independence, but not the reduction of compound lotteries axiom. The subject is assumed to take the BDM procedure as a two-stage lottery. For instance, a subject by setting a minimum selling price of \$4 for a lottery *P*, would define a lottery offering the *P* bet with 40% probability and an uniform distribution lottery between \$4 and \$9.99, $U(4, 9.99)$, with 60% probability, just as in the example above. But instead of reducing this compound lottery to a simple lottery, as above, that is

$(-\$1, \frac{0.4}{36}; \$4, \frac{0.4\times 35}{36}; \$4, \frac{1}{1000}; ...; \$9.99, \frac{1}{1000})$, the subject is assumed to see it as offering the certainty equivalent of P, and the certainty equivalent of $U(4,9.99)$, that is

$[ce_P, 0.4; ce_{U(4,9.99)}, 0.6]$.

A subject maximising a utility function apparently tailor-made for the pair of lotteries above would choose the P bet, ce_P = \$3.86 and $ce_\$$ = \$3.85, but would put a higher price on the $\$$ bet, v_P = \$3.82 and $v_\$$ = \$3.85. One cannot help wondering whether it is possible to find a more convincing example.

The fundamental message from these four arguments is that what looks as preference reversal may not be preference reversal at all. That is, choosing the least valued of two bets need not contradict procedure invariance, and may be accommodated by theories of rational choice. This accommodation comes at a cost. It requires abandoning assumptions – transitivity, independence or the reduction of compound lotteries – that have been regarded as normatively compelling. These assumptions are however not fundamental in a theory of rational choice, whereas the principle of procedure invariance is. Moreover, if one accepts that expectations of regret and rejoicing play a role in rational decisions transitivity should not be assumed, and accommodation of the preference reversal phenomenon by regret theory would come at no cost, except perhaps the cost of some loss of tractability.

One must note though that these four explanations of preference reversal offer only theoretical conjectures that need to be tested empirically.

8. Experiments on the Causes of Reversal

In the two previous sections we saw competing theoretical explanations for preference reversal. This theoretical research motivated a new round of experiments designed to put the several explanations to test.

Loomes, Starmer, and Sugden (1989, 1991) conducted several experiments to test the predictions of regret theory, while controlling for violation of procedure invariance, independence, and of the reduction of compound lotteries axiom. They tested a version o regret theory in which the preference between any two acts x and y with monetary consequences only is given by

$$V(x, y) = \sum_i p_i \Psi(x_i, y_i),$$

where p_i is the probability of event i, x_i and y_i are the consequences of acts x and y in case of event i, and $\Psi(.,.)$ is a skew-symmetric function, that is, $\Psi(a, b) = -\Psi(b, a)$ for any real number a and b. $x \succ y$ iff $V(x, y) > 0$; $x \sim y$ iff $V(x, y) = 0$; and $y \succ x$ iff $V(x, y) < 0$. The version of regret theory presented in the previous section is a special case of the present version. The authors assume *regret aversion*, that is, a disproportionately large aversion to larger regrets. Formally, for any real numbers $a > b > c$, $\Psi(a, c) > \Psi(a, b) + \Psi(b, c)$ (this inequality conveys more readily love for rejoicing, which, given the skew-symmetry of $\Psi(.,.)$, implies regret aversion).

For a certain class of triples of acts, call them *regret triples*, this theory makes predictions that are consistent with observed decision patterns in preference reversal experiments. Table 7 shows a special case of that class. This special case is useful because it relates easily to the tasks in a preference reversal experiment. Table 6, which we used in the previous section to illustrate the connection of regret theory and preference reversal, is a special case of the table below, with the exception that in table 6 $d < e$.

Table 7: Acts in Loomes et al (1989, 1991)

Acts	Probabilities of events		
	p_1	p_2	p_3
$	a	d	d
P	b	b	e
C	c	c	c

$a > b > c > d \geq e$.

Note that $ and P have a structure that is typical of pairs of bets used in the preference reversal experiments. The better outcome of $ is larger than that of P ($a > b$), but the probability of the better outcome is lower in $ than in P ($p_1 < p_1 + p_2$). Some pairs used in experiments verify the condition $d \geq e$, while others do not. Half of the pairs used in Grether and Plott (1979) and in Pommerehne et al (1982) do, and so did all pairs used in Tversky, Kahneman, and Slovic (1990), reviewed below, all of which observed the preference reversal phenomenon.

Loomes et al (1989, 1991) show that, for regret triples, regret theory, with the assumption of regret aversion, is consistent with the intransitive cycle $P \succeq \$$, $\$ \succeq C$, and $C \succeq P$, but rules out the opposite cycle $\$ \succeq P$, $P \succeq C$, and $C \succeq \$$. Note that the former is the choice analogue of a standard reversal, whereas the latter is the analogue of a non-standard reversal. For that reason, we will call them here *standard* and *non-standard* cycles. If $d < e$, both cycles are compatible with regret theory.

The predictions of regret theory depend on the dependence between $\$$ and P. As shown in the table above, the better consequence of $\$$ happens only when the better consequence of P happens, which implies that the worse consequence of P happens only when the worse consequence of $\$$ happens. The authors refer to the dependence as the juxtaposition of the consequences of the acts. The juxtaposition shown in the table above corresponds to a situation in which the winning numbers of the $\$$ bet are a subset of the winning numbers of the P bet, and subjects view the number drawn to determine the outcome of the bet as independent of their choices. It is reasonable to assume the latter condition; the former was met by the pairs used in Grether and Plott (1979), Pommerehne et al (1982), Reily (1982), and Tversky et al (1990).

Loomes et al (1989, 1991) run four experiments involving regret triples.[9] In these experiments 576 different subjects made choices only. For a triple of acts, a subject would make three pairwise choices, one choice from each of ($\$$, P), ($\$$, C), and (P, C). The results were consistent with the predictions of regret theory. Consider an observation a subject's set of three choices over a triple. Aggregating over the four experiments (see Starmer and Sudgen 1998) there were 1856 such observations. Of these, 13.9% were standard intransitive cycles, and only 2.9% were non-standard cycles.

These results mimic the preference reversal phenomenon, but of the several explanations proposed for preference reversal only regret theory can account for this intransitive cycling asymmetry. The cycles were observed in choices only. Therefore they could not result from different mental strategies induced by different response modes. No valuations were elicited, with the BDM or any other procedure, therefore Karni and Safra (1987) and Segal' (1988) conjectures cannot account for these cycles. Holt's (1986) explanation is not valid here either. These experiments used the random lottery procedure. Thus if subjects' preferences do not obey independence their choices may not reflect their preferences in the pairwise choices taken in isolation. For instance, a standard cycle could arise, not because the subject preferred P from (P,

$), *C* from (*P*, *C*), and *$* from (*$*, *C*), but because he preferred the compound lottery (*P*, 1/3; *C*, 1/3; *$*, 1/3) to any other feasible compound lottery. However the same compound lottery arises from a non-standard cycle. Thus, while Holt's (1986) hypothesis accounts for intransitive choice cycles, it does not account for the cycling asymmetry observed in Loomes et al (1989, 1991).

One question that naturally arises is whether the preference reversal phenomenon would be observed with the *P* and *$* bets used in Loomes et al (1989, 1991), and how it would compare with the intransitive cycles and preference reversal observed elsewhere. Most preference reversal studies reviewed so far used a set of pairs of bets originally devised by Lichtenstein and Slovic (1971), or close variations of them. The pairs used in Loomes et al (1989, 1991) are somewhat different, mainly because the probabilities of the better outcome of the *P* bets are lower in Loomes et all (between 60% and 80%) than in most preference reversal experiments (between 81% and 97%).

To answer this question, Loomes et al (1989, experiment 3) compared a choice-only treatment with a typical preference reversal treatment. 186 subjects were randomly allocated to two groups. A first group of 93 subjects made a set of three choices over a triple of acts of the type shown in table 7. Another group of 93 subjects placed monetary values on the *$* and *P* bets of the same triples, and chose between those bets. The designs were perfectly matched: three triples were used, and each of them was dealt with by 31 subjects in each group.

The choices and valuations produced the preference reversal phenomenon, but the asymmetry was lower than usual: out of 93 subjects there were 28 standard reversals and 15 non-standard reversals (each subject dealt only with one pair of bets). The rates of standard and non-standard reversal conditional on choice were 60% and 33%. These become 53% and 15%, similar to those obtained in other studies, if one excludes subjects who gave valuations equal to or higher than the better consequence of the bet, or equal to or lower than the worse consequence.

The other 93 subjects made 14 standard cycles and 4 non standard cycles. The number of cycles and of preference reversals are not comparable, even under the hypothesis that all reversals resulted from intransitive preferences only. The reason is that a subject that prefers *P* to *$*, and whose certainty equivalent of *$* is higher than that of *P* would display a standard reversal in the first group, but would make a

[9] Most observations were obtained from triples that belong to the special case of table 7.

standard choice cycle in the second group only if the monetary value *C* used there happened to fall between his certainty equivalents. Therefore one would expect more standard reversals than standard cycles. To compare both treatments Loomes et al (1989) used the valuations to input the choices between *P* and *C*, and between *$* and *C*, under the assumption of procedure invariance. That is, if a valuation of a bet is higher than *C* a choice of that bet over *C* would be inputted and vice-versa. A valuation equal to *C* was interpreted as equal probability of choosing the bet or *C*.

The choices so imputed produced 11.75 standard cycles and 6 non-standard cycles, not very different from the number of cycles actually observed (14 and 4). Generally the differences between imputed and actual choices was not statistically significant. Therefore there was no evidence that the preference reversals observed in this experiment resulted from violation of procedure invariance.

Starmer and Sugden (1998) showed that the frequency of intransitive cycles seems to depend on how the consequences of the acts are displayed in the choice tasks. Loomes et al (1989, 1991) used a matrix display (see figure 2) in all their experiments. In a series of new experiments Starmer and Sugden used a strip display.

		1 30	31 60	61 100
Matrix display	*$*	£18	£0	£0
	P	£8	£8	£0
		30	30	40

		1 30	31 100
Strip display	*$*	£18	£0
		30	70
		1 60	61 100
	P	£8	£0
		30	40

Figure 2: Matrix and strip displays of a choice task

Both displays show each prospect in a row, and the events that determine the consequence of the acts, in columns. These events are numbers drawn by some random device, and are shown on the top of each column. The number at the bottom of each column is the probability (in percent) of the event. There are two differences between the two displays. The acts are shown together in the matrix display, and separately in the strip display. In the strip display the best outcome of *P* is originated by one event; in the matrix display of the choice between *P* and *$*, that event is split into two.

Starmer and Sugden (1998) ran an experiment with matrix displays that controlled for possible even-splitting effects, and found none. They observed similar frequencies of intransitive cycles both when event-splitting effects were controlled for and when they were not. These frequencies were also similar to those observed in Loomes et al (1989, 1991).

In the experiments that used the strip display, the cycling asymmetry persisted, but the frequency of standard cycles was lower than in the experiments using the matrix display. This is also true when the comparison is restricted to triples that were used with both displays. The authors suggest the following interpretation of these results. People may be motivated by expectations of regret and rejoicing, as assumed by regret theory (or may be using heuristics that can be modelled by that theory). Formation of these expectations requires the comparison of the outcomes of the prospects within events. And the matrix display makes that comparison easier than the strip display. The authors draw two implications from this interpretation. One is that it accepts framing effects, which is a violation of description invariance: the preferences that are revealed depend on how the options are displayed. Another is that these particular framing effect makes sense if one assumes underlying intransitive preferences (that express themselves if the display is favourable). Therefore it is unlikely that the cycling asymmetries observed with the matrix display result from transitive preferences even if one adds some stochastic element.

What is the relevance of this series of experiments for preference reversal? The intransitive choice cycles these experiments uncovered are the choice analogue of preference reversal. The cycles can be explained by regret theory, but not by failure of procedure invariance, or of the independence or reduction of compound lotteries axioms. However it seems that those cycles arise only when the bets are displayed in a way that makes it easy to compare outcomes within events. The preference reversal phenomenon has been observed with bets represented in pie-charts (Grether and Plott 1979, for instance). Comparison of outcomes within events does not seem any easier with pie-charts than with strip displays. This suggests that the behaviour modelled by regret theory cannot be the only factor causing preference reversal. This is of course in line with our conclusion that the valuation of *$* above the winning amount of *P* when *P* is chosen over *$* is incompatible with regret theory. This means that the intransitive choice cycles observed with the matrix display and the preference reversal phenomenon are probably caused by different factors to some extent.

This idea is supported by Tversky, Slovic and Kahneman (1990). These authors ran an experiment that allows the discrimination between intransitivity and violation of procedure invariance, while controlling for violation of the independence and reduction of compound lotteries axioms.

The explanations of preference reversals that are based on the violation of the independence and reduction of compound lotteries axioms involve the BDM procedure in one way or another. To exclude these explanations Tversky et al (1990) modified the usual experiment design by replacing the BDM procedure with what they called an *ordinal payoff scheme*. Subjects stated money equivalents to both bets in each pair and chose one of the bets. Then one pair was selected, and a random device determined whether subjects played the bet they had chosen or the bet they had valued more highly.[10]

Thus in contrast to experiments using the BDM procedure, in which each price a subject sets for a bet defines a different lottery, in this experiment for each pair of bets there are only three possible outcomes: subjects play the *P* bet if they both choose and value the *P* bet more highly; they play the *$* bet if they both choose and value it more highly; or they have equal chances of playing either the *P* bet or the *$* bet if they choose one of the bets and value the other more highly. If preferences obey independence subjects will not desire the latter outcome. If they do not, inconsistencies are still possible, but we would not expect the usual difference between the frequencies of standard and non-standard reversals, since both of them produce the same outcome. One last reason why standard reversals might be more common than non-standard reversals would be the use of a common tie-breaking rule in case subjects wanted to play each of the paired lotteries with 50% probability. But the reason why most people would resort to such a rule is not obvious.

[10] Payoffs were imaginary. But the authors run a control group in which subjects knew that for 15% of them the incentive scheme would actually be applied. The results in this group did not differ from those of the main group, reported here.

As standard reversals were much more frequent than non-standard ones, 45% against 4% of all responses, violations of independence or of the reduction of compound lotteries axiom do not seem to account for them.

To assess the transitivity of preferences Tversky et al (1990) defined for each pair of bets an amount of money, *X*, and asked subjects to choose between the *P* bet and *X*, and between the *$* bet and *X*. About half of the cases of preference reversals met the conditions $P \succ_c \$$ and $v_\$ > X > v_P$. The table below shows the proportions of the four response patterns that meet the above conditions. Of these only 10% were cases of intransitive choices, as table 8 shows. In almost two thirds of the cases subjects chose *X* over the *$* bet, although they had priced the *$* bet over *X*. That is what the authors call overpricing of the *$* bet, noting that they mean overpricing relative to choices, and do not mean the term to imply that valuations are biased and choices reflect true preferences. Underpricing of the *P* bet, and simultaneous overpricing of the *$* bet and underpricing of the *P* bet account for smaller proportions of patterns. Thus the authors conclude that the main cause of preference reversals is a failure of procedure invariance, especially the overpricing of the *$* bet.

Table 8: Distribution of Response Patterns* in Tversky et al. Study 1

Choice Pattern	%*	Diagnosis
$\$ \succ_c X \succ_c P$	10.0	Intransitivity
$X \succ_c P, \$$	65.5	Overpricing of *$*
$P, \$ \succ_c X$	6.1	Underpricing of *P*
$P \succ_c X \succ_c \$$	18.4	Both overpricing of *$* and underpricing of *P*

* % of the 620 cases that met the conditions $P \succ_c \$$ and $v_\$>X>v_P$.
Source: Tversky, Slovic and Kahneman (1990).

The authors run a second experiment with the same design but involving delayed payments, all of them imaginary. In each pair there was a short term option, offering a relatively small amount of money after a relatively short period of time, and a long term option, offering a bigger amount more delayed in time. The results were qualitatively similar. Subjects quite often chose the short-term option over the long term one, but priced the long-term option above the short term one. As in the previous study reversals were due mainly to overpricing of the long-term option, with intransitivity and the other forms of mispricing playing a smaller role.

Tversky et al see their results as confirming the scale-compatibility hypothesis, the idea that decisions are mostly influenced by the dimension most compatible with the response mode. Choices either between lotteries or between delayed payments are not particularly more compatible with one dimension than the other, probability and payoff in lotteries, time and payoff in delayed payments. But valuations are more compatible with payoffs in both cases. Therefore the scale-compatibility hypothesis predicts payoffs to have a bigger influence in valuations of both bets and delayed payments than in choices. Hence the overpricing of *$* bets and long term payments.

The authors also suggested the *prominence hypothesis* as an explanation. We saw in section 6 that a likely cause of the prominence effect is the strategy-compatibility hypothesis. Tversky et al (1990) and the study reviewed below work with the prominence hypothesis. This merely states that the prominent attribute is given more weight in choices than in matching or valuations, without implying any particular explanation, such as strategy compatibility. If probabilities are the more prominent dimension in bets and time the more prominent dimension in delayed payments the prominence hypothesis could explain the observed results too.

Cubitt, Munro, and Starmer (forthcoming) run an experiment to discriminate between the scale-compatibility and the prominence hypotheses. One group of subjects chose between pairs of bets and valued the bets according to Tversky et al's (1990) ordinal payoff scheme. The scheme was applied, and subjects played a bet for real at the end of the experiment. The results obtained were qualitatively similar to those observed by Tversky et al (1990).

Another group of subjects performed the same choice tasks, but valued the bets in terms of probabilities. Each bet to be valued, with expected values ranging from £2.50 to £9.90, was presented next to a "yardstick" bet offering £10 with an unspecified probability. Subjects were asked to specify the probability so that they would be indifferent between both bets. They knew that at the end of the experiment a pair of bets would be selected, and that they would play with equal probability either the one they had chosen or the one they had given the highest probability valuation.

Whereas with monetary valuations both the sacale-compatibility and the prominence hypotheses work in the same direction, with probability valuations they work in opposite ways. The prominence hypothesis predicts the usual kind of reversals, since according to this hypothesis the reversal is caused by the probabilities having more influence in choices than in valuations, regardless of the type of valuation. The compatibility hypothesis predicts the overvaluation of *P* bets, thus leading to the less common type of reversal. In this group of subjects both types of reversals were about equally common.

One possible explanation for the results of both groups is that both hypotheses are correct. If so, in the probability-valuation group the probability could have been given similar weights in choices and valuations: in choices because, of prominence; in valuations, because of scale compatibility. In that case a similar frequency of both types of reversal could result from random errors.

What do the experiments reviewed in this section tell us? Violation of the independence or the reduction of compound lotteries axioms are not the cause of preference reversal. Systematic intransitivities arise in some circumstances, but they seem to play only a minor role in preference reversal. That leaves us with violation of procedure invariance as the likely explanation. Scale compatibility and strategy compatibility have been proposed as explanations of preference reversal and the related choice-matching discrepancy. Cubitt et al's (forthcoming) results suggest that neither hypothesis on its own can explain preference reversal, but the two together can. Yet the phenomenon is still puzzling, especially in its simple form uncovered by Tversky et al (1990). Here many subjects valued a *$* bet at a certain amount, and chose a lower amount over that bet.

The literature on preference reversal reviewed so far provides a good illustration of the methods of an experimental science: a phenomenon was observed. theoretical work produced explanatory hypotheses; and experiments were designed to test those hypotheses. The experiments elicited decisions in controlled conditions. These conditions were set up so that the hypotheses yielded unambiguous predictions concerning the decisions. Finally the hypotheses were tested by comparing predicted and observed decisions.

This method proved fruitful, but we could do more. We could try to observe the decision process. That is what I will try to do in chapter 2. Observation of decision and other mental processes can be attempted by means of think-aloud experiments. In a typical think-aloud experiment subjects are asked to speak aloud their thoughts while performing a task. This method has widespread use in psychology, and it is surprising that it has not been used to study preference reversal.

Economists in particular are likely to feel sceptical about the usefulness of asking subjects to speak aloud their thoughts. What incentives does a subject have to say what is really going on in his mind? The experiment reported in chapter 2 addresses that concern: it is incentive compatible, that is, the design makes it the subjects' interest to talk sensibly. This is an innovation relative to the typical think-aloud design.

The think-aloud method can be used to test hypotheses, as long as those hypotheses yield predictions as to what people will say. For instance, if, as hypothesised by Holt (1986), a subject views her decisions in the whole experiment as determining a compound lottery, that should show up in what she says. But an advantage of the think-aloud method is that it may also suggest hypotheses. For instance, if subjects in a think-aloud experiment behaved as hypothesised by Holt, the experimenter might notice it, even if Holt had not formulated his hypothesis yet.

The main aim of chapter 2 is methodological, but one cannot rule out the possibility of learning something about preference reversals that we do know yet.

9. Experience, Markets, and the Real World

A number of authors claim that preference reversal has been observed only with inexperienced subjects in environments that do not resemble markets in the world outside the laboratory. Some of these critics elicited preferences in experimental settings designed to be more representative of actual markets. Some allowed subjects to acquire experience through repetition and feedback. All of them observed a subsidence or disappearance of preference reversal. Here follows a review of five of these studies.

Bohm (1994a) questions the relevance of the typical preference reversal experiment to economics. He argues that the lotteries subjects faced in those experiments are not traded in the "real-world," that the monetary values involved were insignificant, and that subjects did not seek to make decisions on those lotteries.

Instead he elicited preferences on used cars, from subjects who were willing to buy them. The cars were a 1976 Volvo and a 1977 Opel, both with mileages over 180,000km. The author argues that such cars are lotteries, in the sense that their future performance is uncertain. Furthermore the Volvo, seen as a reliable but dull car by young Swedes, the group to which subjects in this experiment belonged, could be considered a *P* bet; whereas the Opel, with its "theatre-red plush upholstery and a good radio," could be seen as a *$* bet: more likely to break down than the Volvo, but more enjoyable to drive if it did not.

Subjects were 26 students that registered as potential buyers for the cars. They were asked to state their preference between the cars and to make their bids to buy each of the cars. To motivate subjects to state their true preference in the choice task one of them was selected at random to receive one of the cars for free: his preferred car with 2/3 probability, his least preferred car with 1/3 probability. Another lottery determined whether the winner would keep the car or receive his bid for that car instead. This was intended to motivate subjects to make considerate bids for both cars. Finally one or both cars would be sold to the highest bidder for a price equal to the second highest bid. The buyer of the first car would be excluded from the second auction, if there was one.

22 of the 26 subjects made bids strictly in line with their choices. The remaining four made equal bids, but did not state indifference, which they were allowed to do. The author concludes that his study shows that preference reversal cannot, without further evidence, be assumed to occur outside the laboratory.

This is a highly contestable conclusion. It is a stretch to take the two used cars as *P* and *$* bets. Specifically it is not clear that they possess the features that psychologists suggest are the cause of preference reversal.

According to the prominence hypothesis people choose from close alternatives the one that is superior on the prominent attribute. This hypothesis predicts preference reversal in gambling decisions if the prominent attribute of a gamble is the probability. This could predict preference reversal in Bohm's (1994a) study if the author's account of the Volvo and Opel as a *P* and *$* bets is correct, and if reliability is the prominent attribute in used cars. None of this is obviously true. Even if probability is the prominent dimension in bets it does not follow that reliability is the prominent dimension in used cars. Besides the prominence hypothesis cannot by itself explain preference reversal, as it predicts patterns of reversal that are not observed when probability valuations are used (Cubbit et al, forthcoming).

The other psychological explanation for preference reversal is based on a overvaluation of the *$* bet: this would be caused by the compatibility between the amount to win and the monetary valuation. This hypothesis, even accepting Bohm's (1994a) account of the used cars as pair of *P* and *$* bets, does not predict a similar overvaluation of the Opel, the presumed *$* bet in this study. According to the author, the high prize in this car would comes in the form of pleasurable driving, which is not more compatible with the monetary valuation than the reliability of the Volvo. Nor is it clear that any of the cars might be obviously superior to the other in any other dimension that is particularly compatible with the monetary valuation.

Furthermore, subjects faced only one pair of options, and had five days to consider their choices and bids. This feature alone might greatly decrease preference reversal in typical gambling decisions. Subjects might well perceive a logical connection between the choice and the bids, and let their decisions be influenced by consistency considerations. This would be even more likely to happen if subjects faced some probability of receiving the amount they bid for their chosen option. This feature of Bohm's (1994a) design helps further the perception of a logical connection between choices and bids.

The disappearance of preference reversal in gambling decisions in such a design would not mean the irrelevance of the phenomenon, as many decisions one has to make take a single form, say, choice or valuation; one is not usually required to ponder all these forms of decisions together when regarding a set of options.

Bohm (1994b) ran an experiment with delayed payments along the lines of Tversky et al (1990). All treatments used a short term option offering SEK 1000 (1000 Swedish kronor, about USD 200 at 1992 exchange rates) after three months, and a long term option offering SEK 1200 after fifteen months. One of the treatments included three additional pairs of options. Subjects were third year students specialising in finance at two top Swedish schools of Economics, and middle rank bank employees.

They were asked to state their choice, and bids to buy in a second-price auction. For a group of 32 students the auction was for real. The students were divided in three groups, and the highest bidders would actually buy the claim (if the same person was the highest bidder for both claims, he would buy only the one he had chosen). The choice would be for real only with probability 10%. For all the other subjects the claims were hypothetical. Additionally all subjects were asked to state minimum selling prices, but again, these would have no real consequences.

The groups making hypothetical decisions produced an asymmetric pattern of reversal similar to that observed by Tversky et al (1990), both with buying and selling prices: among the subjects that chose the short-term claim, between 62% and 81%

valued the long-term claim more highly; among the subjects that chose the long-term claim, less than 15% valued the short-term claim more highly.

In the group facing real claims, the comparison of bids to buy and choices gives rise to relatively few and roughly symmetric reversals: 15% of reversals among subjects that chose the short-term claim, and 18% of reversals among the subjects that chose the long-term claim. These inconsistencies could possibly be the result of some unbiased random element in decision making.

Tversky et al (1990) showed that preference reversal in decisions concerning hypothetical delayed payments occurs among inexperienced subjects. Bohm (1994b) shows that the phenomenon persists even among subjects familiar with financial matters if payments are hypothetical. But the phenomenon may no longer occur when familiarity and real payments are combined.

It would be interesting to know whether financial sophistication is necessary or whether real payments are enough to reduce, maybe eliminate, preference reversal in decisions concerning delayed payments.

Berg et al (1985) investigated the effects of arbitrage and experience in gambling decisions. The experiment had two identical runs. In one treatment each run replicated Grether and Plott's (1979) experiment 1 with incentives: subjects would face six pairs of bets, choose one from each pair, and value all of them according to the BDM procedure. At the end of the first run, one task was selected at random and was played out for real. Then subjects would proceed to the second run.

Another treatment included all this and added a money pump. Subjects were not informed of the money pump, but were told that the experimenter was free to make any transactions with them that were in accordance with their decisions. At the end of the eighteen tasks subjects would go through one round of arbitrage for every pair in which they had reversed. Suppose that in a pair a subject chose *P* over *$* and valued *P* at $3 and *$* at $5. The experimenter, after the eighteen tasks had been performed, would sell him the *$* bet for $5, give him the *P* bet in exchange for the *$* bet, and finally would buy back the *P* bet for $3. The subject would loose $2 from his initial stake of $7. All rounds of arbitrage were for real. After the arbitrage subjects would play one of their decisions for real. Then subjects would proceed to a second, identical run.

The maximum price a subject is willing to pay for a bet may be lower than the minimum price he demands for the same bet. The BDM instructs subjects to state the latter. A rational subject might have to ignore that instruction. To address this issue two other treatments, with and without arbitrage, used a different price-elicitation procedure that gave subjects no conflicting instructions. The results obtained under each price-elicitation procedure were basically the same, therefore I will report the results aggregated across pricing schemes.

The breakdown between standard and non-standard reversals is reported only for the replication of Grether and Plott (1979), that is, the BDM treatment without arbitrage. The reversal pattern was asymmetric but less so than in Grether and Plott: the rates of standard and non-standard reversal conditional on choices were respectively 0.51 and 0.24 in the first run, and 0.58 and 0.16 in the second run.

Table 9 summarises the main results. Experience seemed to have reduced both the frequency, especially in the arbitrage group, and the dollar magnitude of reversals. The magnitude of a reversal is the difference between the prices stated for two paired bets when a reversal occurs.

Table 9: Berg et al (1985): effects of arbitrage and experience

	Mean reversals per subject		Mean magnitude of reversals*	
	1st run	2nd run	1st run	2nd run
No arbitrage	1.89	1.68	4.79	3.41
Arbitrage	2.42	1.60	3.27	1.77

* Difference in US dollars between the values placed on two paired bets when a reversal occurred. For comparison, expected values ranged from USD 1 to USD 4.
Source: Berg et al (1985), tables 5 and 6.

The obligation by subjects to act according to their decisions if the experimenter so wished had unclear effects in the first run. The number of reversals was higher in the arbitrage than in the no-arbitrage treatment, which suggests that subjects did not expect the money pump and were not concerned with the possibility of having to pay their stated prices to buy additional bets. On the other hand, the mean magnitude of reversals was lower in the arbitrage treatment, which suggests the opposite. There is a possible explanation for this puzzle. Some subjects may have priced *$* bets low, fearing having to buy them, making many non-standard reversals. These are usually of small magnitude, because the value placed on the *P* bets is usually low. The published data is insufficient to test this hypothesis.

If the possibility of arbitrage went unnoticed or failed to have a clear effect, being exploited by arbitrage seemed to have a clear effect. There was a large reduction in the number of reversals from the first to the second run in the arbitrage group, but hardly any reduction in the no-arbitrage group. This is the main, if unsurprising, conclusion to draw from this study: after being exploited by a money pump, subjects become less inconsistent.

Chu and Chu (1990) "purified" Berg et al's (1985) design, and made preference reversal disappear altogether. In this experiment no bets were played or sold in a BDM procedure. All that was left was the money pump. Subjects would be individually presented a pair of bets at a time, and would state their preferences and "fair" prices for the bets. They knew that if asked to they would have to make any transactions that were according to their stated preferences and prices. Then if prices were consistent with expressed preferences nothing would happen, and the subject would go on for the next pair of bets; otherwise the subject would go through one round of arbitrage, as in Berg et al (1985). Then the whole procedure would be repeated with the same pair of bets as many times as it would take for prices to be consistent with expressed preferences, after which the subjects would be presented the next pair of bets. All rounds of arbitrage counted for real.

As one would expect from such a pure money-pump design, subjects learned quite quickly to be consistent. The subjects who reversed with the first pair of bets, 29 out of 83, needed on average to go through 1.7 rounds of arbitrage to learn to avoid the money pump. For most of them the lesson was not forgotten when they proceeded to the following pairs of bets: only 5 reversed with the second pair, and none with the third. Of the 54 subjects who did not reverse in the first pair 11 reversed in the second. Again, of these 11, only 1 reversed with the third and last pair of bets. Of the 43 who did not reverse neither in the first nor in the second pairs 6 reversed in the third. The authors conclude that in a simplified market-like environment where inconsistencies are exploited by arbitrage preference reversal diminishes, and that the learning effect is lasting.

Arbitrage has long been recognised as a major force conditioning agents that are primarily interested in market values, such as traders in currency and financial markets. It is not clear that it may play a similar role in decisions that are primarily determined by subjective value, such as consumers' decisions. This is of course the area where preference reversal has been observed. It does not seem that individuals acting on subjective value in ordinary markets will see possible biases in their decisions exposed and punished by such a transparent mechanism as the money pump. Note also that in this study subjects' preferences regarding the lotteries were irrelevant, as subjects stood no chance of playing them. Thus this study does not provide evidence that market environments will diminish preference reversal.

Cox and Grether (1996) run a preference reversal experiment in which valuations were elicited by means of selling auctions. Some subjects in groups of five participated in a second price auction (SPA): they submitted sealed bids to sell a bet, and the experimenter would buy the bet from the lowest bidder at the second lowest price; the other four subjects played the bet. Other subjects participated in an English clock auction (ECA): the auctioneer kept reducing the price until all subjects withdrew from the auction; the last to withdraw sold the bet for the price at which the previous subject had withdrawn, the others played the bet. Each bet was auctioned five times in a row. Another group of subjects valued the bets by means of the BDM procedure, also five times in a row. After both bets were priced subjects chose one of them. Subjects knew the results of their decisions, including their accumulated earnings as the experiment went on. Reversal rates were computed by comparing the single choice with the first and fifth valuations, and a summary of the results is shown in the table 10.

There is a decrease of the overall reversal rate between the first and fifth valuations in all treatments. The striking result is however the disappearance of the usual asymmetry between standard and non-standard reversals in the two markets where decisions were made for real money. In the English-clock auction the opposite asymmetry seems to emerge. However the 63.6% rate of non-standard reversal obtained with the fifth valuations refers to only seven non-standard reversals out of eleven *$* choices, thus one cannot give much importance to this result.[11]

[11] The changes between the first and last valuations may actually be understated by the authors' methodology. Besides valuing bets of a pair five times by bidding at an auction subjects also valued bets of other pair once using the BDM procedure. Subjects made these 12 valuations and two choices, ch(.), in one of the following orders:

$P_1\ P_1\ P_1\ P_1\ P_1 \quad \$_1\ \$_1\ \$_1\ \$_1\ \$_1 \quad \text{ch}(P_1, \$_1) \quad P_2\ \$_2 \quad \text{ch}(P_2, \$_2)$;

$P_2\ \$_2 \quad \text{ch}(P_2, \$_2) \quad P_1\ P_1\ P_1\ P_1\ P_1 \quad \$_1\ \$_1\ \$_1\ \$_1\ \$_1 \quad \text{ch}(P_1, \$_1)$;

$\$_1\ \$_1\ \$_1\ \$_1\ \$_1 \quad P_1\ P_1\ P_1\ P_1\ P_1 \quad \text{ch}(P_1, \$_1) \quad \$_2\ P_2 \quad \text{ch}(P_2, \$_2)$;

Table 10: Rates of reversal conditional on choices (%) in Cox and Grether (1996)

	N	Choice of *P* bet, %	1st valuation		5th valuation	
			SR	NSR	SR	NSR
Second-price auction: incentives	60	60.0	75.0	8.3	27.8	33.3
fixed pay	20	55.0	36.4	44.4	21.3	33.3
English-clock auction: incentives	40	72.5	62.1	72.7	31.0	63.6
fixed pay	20	70.0	92.9	16.7	71.4	16.7
BDM: incentives	20	50.0	72.7*	11.1*	40.0	0.0

Key: N, number of choices; SR and NSR, standard and non-standard reversals; subjects with incentives received the full amount of the outcomes of their tasks or half that amount.
* These are the reversal rates on 60 choices, not 20.
Source: Cox and Grether (1996), table 1, and text.

The behaviour of participants in the English clock auction presents two other peculiarities. Subjects with incentives exhibited similar rates of standard and non-standard reversals right from the start. This lends support to psychological explanations of reversals, since in this auction subjects keep choosing between the bet and the current price. Subjects who received a fixed payment displayed the asymmetrical pattern of reversals in both first and last valuations. The authors offer the following explanation for this. Subjects may have become bored of waiting for the price to fall, and as their earnings did not depend on their decisions they may have withdrawn from the auction at an early stage, when the price was still above the minimum they would actually be willing to accept. This would affect *$* bets much more than *P* bets. The price in the auction started at the positive outcome of the bet, and decreased five cents a second. So for instance in the case of the pair of our examples, one of the two pairs of bets used in this experiment, it would take four minutes for the price of the *$* bet to decrease from $16, the winning prize of the *$* bet, to $4, the starting price of the *P* bet. For the *$* bet of the other pair the corresponding time was two minutes and twenty seconds.

The reduction of the overall reversal rate and the disappearance of the usual asymmetry in the markets with incentives leaves open the possibility that the

$\$_2$ P_2 ch(P_2, $\$_2$) $\$_1$ $\$_1$ $\$_1$ $\$_1$ $\$_1$ P_1 P_1 P_1 P_1 P_1 ch(P_1, $\$_1$).

This means that the first market valuation of P_1 for instance was actually an average of the first, third, sixth and eighth of all valuations. And it is possible that when bidding to sell P_1 subjects may have benefited from the experience acquired with other bets.

On the other hand the experience from valuations may also have affected the choices. Comparison of first valuations with an initial choice might have produced different reversal rates.

inconsistencies are due to random errors rather than the kind of response-mode-induced bias put forth by psychologists. The authors attribute this reductions or possible disappearance of preference reversal to repetition, feedback, and market experience. Repetition and feedback seem not to have been enough, since the asymmetry remained in the BDM group.

However it is not clear how market experience might achieve this. The authors found bids in auctions to be positively correlated with the last market price. This contributes to drive down the bids for the *$* bets, as they tend to be very dispersed and the market price is the second lowest bid. The *P* bets should be much less affected, as the dispersion of their valuations is usually very low. This contributes to a more symmetrical pattern of reversals. But it raises the question of whether the last valuations reflect subjects' preferences any better than the first ones, as it is hard to see how the market prices should help subjects discover their preferences. Instead subjects might have just engaged in a sort of competition to sell the bets.

The role of repetition and feedback is easier to understand than that of market experience from the point of view of preference theory. After having played the bet a couple of times, subjects should be in a better position to know how much the bet was worth for them. Regressions estimated by the authors indicate that the loss component of bets did not influence the first bid, but influenced the subsequent ones. Attention to this component may have been drawn by losses, quite likely in the case of *$* bets, in the early sessions of the auction.

Subjects were more likely to play the bets in the auctions than in the BDM treatment. In the auctions four out of five subjects played their bets in each session. In BDM treatment the counter-offer was uniformly distributed between $0.00 and $9.99. Thus it would take a valuation of $8 for a subject to face the as high a probability of playing the bet as subjects faced ex ante in the auctions. A valuation of $8 for the bets used in this experiment, with expected values between $1.35 and $3.86, would denote an extraordinary love for risk.

If playing the bets, and presumably experiencing losses in the case of the *$* bets, was an important factor in driving the *$* bids down, and if *$* bets were played less often in the BDM treatment than in the auctions, we would expect the reduction of the typical, asymmetrical reversal pattern to be larger in the auctions than in the BDM treatment. We would then have an alternative explanation for the results of Cox and Grether (1996), and one that is more readily understandable from the point of view of preference theory than the authors' hypothesis that the market played a special role beyond repetition and feedback.

These five studies show the subsidence or disappearance of preference reversal in particular market environments. Bohm (1994a), and Chu and Chu (1990) offer little evidence that market experience will reduce or eliminate preference reversal in general. One would not necessarily expect preference reversal to occur in decisions concerning used cars, or to persist under the transparent exploitation by a money pump. This does not mean that preference reversal will not persist in other market environments.

Berg et al (1985) offer some evidence that experience may reduce preference reversal, even without the market environment: experience seemed to have reduced the magnitude of reversals in the no-arbitrage group, even if the number of reversals barely decreased.

Bohm (1994b) and Cox and Grether (1996) make a more credible claim that market experience will reduce or even eliminate preference reversal. Both studies look for preference reversal in circumstances where they had previously been found. The only major differences relative to previous experiments was the use of markets to elicit valuations, and the use of experienced subjects (in Cox and Grether 1996 experience was acquired during the experiment). These changes were enough to reduce the incidence of preference reversal. More, the disappearance of the reversal asymmetry suggest that the inconsistencies that remained may result from randomness alone.

These studies raise a number of questions. What were the factors behind the reduction of standard reversals? Was it simply experience? That is, was it just the case that Bohm's (1994b) subjects were acquainted with financial matters, and Cox and Grether's (1996) subjects learnt how much they subjectively valued the lotteries by playing them? Or did the markets play a special role? For instance, are valuations in markets, as Cox and Grether (1996) suggest they may be, psychologically different tasks from the same valuations in non-market environments? If so, does the market environment promote rational decisions, or does it simply induce a different sort of bias? Did subjects extract useful information from the market price? (This could obviously not apply to the one-shot Bohm's (1994b) experiment.) Or did they mechanistically adjusted their bids towards it?

Chapter 3 reports an experiment that addresses these questions. One treatment basically replicates Cox and Grether's (1996) second-price auction. In another treatment valuations are elicited in a second-to-last price auction. If markets generally promote rational decisions, we should observe the reduction or disappearance of preference reversal in both treatments. In contrast, the alternative explanations alluded to above for Cox and Grether's (1996) results predict that preference reversal will persist or even become stronger in the second-to-last price auction.

10. Are Preferences Discovered or Constructed?

Cox and Grether's (1996) results are used by Plott (1996) to illustrate his *discovered preference hypothesis*. As he claims, this hypothesis is not a theory from which quantifiable predictions may be derived, but a way of interpreting results of economists and psychologists' experiments. This hypothesis maintains that people have stable preferences, but their actions may fail to reflect those preferences when people are facing *new tasks*, or in complex situations requiring anticipation of *other agents*' rationality. Rationality then evolves through three stages. In the first stage the individuals are inexperienced about the environment and about the consequences of their actions, and their behaviour may seem erratic, or even exhibit systematic irrational features. Stage two occurs when people have through practice and feedback learnt to deal with the new tasks, that is, when the tasks are no longer new. The evolution of rationality culminates in stage three, when people have learnt to anticipate the other agents' rationality, with the discovery of stable and consistent preferences. According to this hypothesis "attitudes like expectations, beliefs, risk aversion and the like are *discovered*" (p. 227). Social institutions play an important role in achieving this rationality.

With this hypothesis Plott tries to conciliate "the power of models built on principles of rational choice, or on related concepts of purposeful choices, to predict the behaviour of groups of people, such as committees and markets" (p. 226) with the anomalies found in experiments on individual decision making. The anomalies, the author maintains, are caused by the inexperience of the first stage of the evolution of rationality. And he reviews four examples of such anomalies. The conclusion he draws from Cox and Grether's (1996) experiment is that "the classical preference reversal can be seen as a product of inexperience and lack of motivation, and it goes away with experience in a market setting" (p. 231).

The discovered preference hypothesis as a way of accommodating within a broader model of rational choice the anomalies revealed by experimental evidence was developed at least partly in opposition to the psychologists' idea of constructed preferences. This idea of constructed preferences encompasses the compatibility and prominence hypotheses. It assumes that people have various motivations, and when they face choices between objects with several attributes these motivations may conflict.

The theory of rational choice assumes people are able to make trade-offs, and arrive at a global comparison of the objects. Instead, argue the psychologists, people compare an attribute of an object with the corresponding attribute of the other object, and, depending on how the problem is framed, and on a number of other contextual factors, give more weight to different attributes, thus arriving at different preference orders. Another idea is that people try to change the way a problem is viewed in ways that make the justification for the decision they are about to make appear more compelling. In this sense preferences are not discovered, or do not even exist prior to the decision task, but are constructed in the process of decision. Slovic (1995, p. 369) writes: "Construction strategies include anchoring and adjustment, relying on the prominent dimension, eliminating common elements, discarding nonessential differences, adding new attributes into the problem frame in order to bolster one alternative, or otherwise restructuring the decision problem to create dominance and thus reduce conflict and indecision. As a result of these mental gymnastics, decision making is a highly contingent form of information processing, sensitive to task complexity, time pressure, response mode, framing, reference points, and numerous other contextual factors."

11. Does Preference Reversal Matter?

What do we know now? Preference reversal has been observed in decisions on gambles, delayed payments, and a variety of other objects. The phenomenon has been studied more thoroughly in gambling decisions than in other areas. In gambling decisions there seems to be a difference in behaviour between experienced and inexperienced subjects. Inexperienced subjects in one-shot experiments produce a highly asymmetric reversal pattern, and there is mounting evidence that the main explanation for that behaviour is violation of procedure invariance. Subjects with previous experience (Bohm 1994b) or that have the opportunity to acquire experience through repetition and feedback during the experiment (Cox and Grether 1996) produce fewer and more symmetrical inconsistencies. This behaviour could possibly be accommodated by stochastic versions of traditional economic theories of rational choice.

This thought assumes that a stochastic theory of rational choice cannot accommodate an asymmetric pattern of reversals. We cannot be sure of whether it can or not before we try. There is however a reason to suspect that it cannot. Suppose that a stochastic theory of rational choice is found that accounts for asymmetrical reversal patterns. Will it then also account for symmetric patterns, or for the disappearance of reversals? This is a problem that any theory aiming to explain behaviour in preference reversal experiments must tackle. It must decide whether it wants to explain behaviour in the one-shot experiment or the behaviour after it has stabilised after enough repetition and feedback. Otherwise a theory could try to account for the changing behaviour, but that would not be a typical economic rational choice theory, which assumes stable preferences.

Assuming that the symmetric reversals can be accommodated by a stochastic theory of rational choice, the difference between the behaviour of inexperienced and experienced subjects fits nicely into Plott's (1996) discovered preference hypothesis: subjects begin to act as if according to a single set of preferences after a period of learning. However Plott and Cox and Grether attribute to markets a role in the learning process that goes beyond repetition and feedback. This special role is not easily understandable. And if markets play such a special role, will they foster rationality or simply induce a different sort of bias?

Another question is whether actual markets provide the same sort of repetition and feedback as Cox and Grether's (1996) simple market did. The answer is yes when it comes to goods that one buys frequently, such as newspapers or coffee, but is less clear when it comes to decisions one makes infrequently, such as buying a house, or choosing a job. On the other hand, in other aspects actual markets provide a better environment for making rational choices than experimental markets. In an actual market one has time to ponder one's decisions, whereas in an experimental market or other economics experiments one often has less than an hour to understand fairly complex environments, and make over a dozen decisions.

Does it matter whether preference reversal can be accommodated by a theory of rational choice? Does it matter whether in markets people act on stable preferences or on constructed preferences?

Most economists would probably agree that the issue is at the very least a matter of intellectual interest. Beyond that there may be disagreement. One could take the following view. The aim of economics is to predict market and other social processes outcomes. If models based on the assumption of individual rationality are able to make correct predictions, it does not matter whether the assumption is descriptively correct.

I do not agree with this position. The assumption of rationality is not merely a useful approximation that enables us to predict market outcomes. It underlies value judgements of those market outcomes. The concept of Pareto efficiency, which is used to make value judgements, is based on the assumption of rationality. For instance, if preferences are constructed is there still an efficiency argument against distorting policies? This thesis will not try to answer this question, but it seems an issue worth pondering.

An area in which preference reversals should be of immediate practical consequence is the use of contingent valuations to inform public policy. Contingent valuations resemble experiments with hypothetical outcomes. If preferences revealed in valuations differ from those revealed in choices, relying on contingent valuations may lead to under-funding or over-funding of policy programs. This is only true, of course, if we still have a criterion to determine the proper level of funding.

To conclude, preference reversal matters, and this overview identified some open questions that the remainder of this thesis will address. Chapter 2 explores the potential of the think-aloud methodology in the study of individual decision making. Additionally, the observation of subject's decision process may suggest new explanations of preference reversal. Chapter 3 tries to clarify the effects of markets and feedback on preference reversal. Of particular interest here is whether market experience will in general reduce, or even eliminate preference reversal. Chapter 4 develops a stochastic model of rational choice and valuation, and fits it to preference reversal data. Such models are necessary, because without them we cannot be sure about what patterns of reversal constitute a non-random deviation from rational behaviour. The concluding chapter brings together the contributions of this thesis to our understanding of preference reversal.

References

Becker, G. M., M. H. DeGroot, and J. Marshak (1963), "Measuring Utility by a Single-Response Sequential Method," *Behavioural Science* 9, 226-32**.**

Berg, Joyce, John Dickhaut, and John O'Brien (1985), "Preference Reversal and Arbitrage," in Vernon Smith (ed.) *Research in Experimental Economics* 3, JAI Press, 31-72.

Bohm, Peter (1994a), "Behaviour under Uncertainty without Preference Reversal: A Field Experiment," *Empirical Economics* 19, 185-200.

–––––– (1994b)**,** "Time Preference and Preference Reversal among Experienced Subjects: The Effects of Real Payments," *The Economic Journal* 104, 1370-78.

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

Chu, Yun-Peng, and Ruey-Ling Chu (1990), "The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note," *American Economic Review* 80, 902-11.

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

Cubitt, Robin, Alistair Munro and Chris Starmer (forthcoming), *Preference Reversals: An Experimental Investigation of Economic and Psychological Hypotheses,* The Economic Journal.

Davis, Douglas D. and Charles A. Holt (1993), *Experimental Economics*, Princeton, Princeton, University Press.

Fischer, Gregory W., and Scott Hawkins (1993), "Strategy Compatibility, Scale Compatibility, and the Prominence Effect," *Journal of Experimental Psychology: Human Perception and Performance*," Vol. 19 No. 3, 580-97.

Fischer, Gregory W., Zvi Carmon, Dan Ariely, and Gal Zauberman (1999), "Goal-based Construction of Preferences: Task Goals and the Prominence Effect," *Management Science*, Vol. 45, No. 8, 1057-75.

Grether, David M. and Charles R. Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69, 623-38.

––––––**, and –––––– (1982)**, "Economic theory of Choice and the Preference Reversal Phenomenon: Reply" [to Pommerehne et al. (1982)], *American Economic Review* 72, 575.

Harless, David H. and Colin F. Camerer (1994), "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica* 62, 1251-89.

Harrison, Glenn W. (1992), "Theory and Misbehaviour of First-Price Auctions: Reply," *American Economic Review* 82, 1426-43.

–––––– **(1994)**, "Expected Utility Theory and the Experimentalists," *Empirical Economics* 19, 223-53.

Hey, John D. and Chris Orme (1994), "Investigating Generalisations of Expected Utility Theory Using Experimental Data," *Econometrica* 62 1291-326.

Holt, Charles (1986), "Preference Reversals and the Independence Axiom," *American Economic Review* 76, 508-15.

Kahneman, Daniel (1996), "Comment" [on Plott (1996)], in K. Arrow, E. Colombatto, M. Perlman, and C. Schmidt, editors, *The Rational Foundations of Economic Behaviour*, Macmillan.

Karni, Edi, and Zvi Safra (1987), "'Preference Reversals' and the Observability of Preferences by Experimental Methods," *Econometrica* 55, 675-85.

Lichtenstein, Sarah and Paul Slovic (1971), "Reversals of Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89, 46-55.

––––––, **and ––––––** **(1973)**, "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas," *Journal of Experimental Psychology*, 101, 16-20.

Lindman, Harold R. (1971), "Inconsistent Preferences among Gambles," *Journal of Experimental Psychology* 89, 390-97.

Loomes, Graham, Chris Starmer, and Robert Sugden (1989), "Preference Reversal: Information-Processing Effect or Rational Non-Transitive Choice?" *The Economic Journal* 99, 140-51.

––––––, ––––––, **and ––––––** **(1991)**, "Observing Violations of Transitivity by Experimental Methods," *Econometrica*, Vol. 59, No. 2, 425-39.

Loomes, Graham and Robert Sugden (1983), "A Rationale for Preference Reversal," *American Economic Review* 73, 428-32.

––––––, **and ––––––** **(1995)**, "Incorporating a stochastic element into decision theories," *European Economic Review* 39, 641-48.

Plott, Charles R. (1996), "Rational Individual Behaviour in Markets and Social Choice Processes: the Discovered Preference Hypothesis," in K. Arrow, E. Colombatto,

M. Perlman, and C. Schmidt, editors, *The Rational Foundations of Economic Behaviour*, Macmillan.

Pommerehne, Werner W., Friedrich Schneider and Peter Zweifel (1982), "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination," *American Economic Review* 72, 576-74.

Reilly, Robert (1982), "Preference Reversal: Further Evidence and Some Suggested Modifications in Experiment Design," *American Economic Review* 72, 577-84.

Roth, Alvin E. (1995), "Introduction to Experimental Economics" in John H. Kagel and Alvin E. Roth, editors, *Handbook of Experimental Economics*, Princeton, Princeton University Press.

–––––– **(1996),** "Individual Rationality as a Useful Approximation: Comments on Tversky's 'Rational Theory and Constructive Choice'," in K. Arrow, E. Colombatto, M. Perlman, and C. Schmidt, editors, *The Rational Foundations of Economic Behaviour*, Macmillan.

Segal, Uzi (1988), "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review*, March 78, 232-36.

Seidl, Christian (2002), "Preference Reversal," *Journal of Economic Surveys* 16, 621-56.

Slovic, Paul (1975), "Choice between Equally Valued Alternatives," *Journal of Experimental Psychology, Human Perception and Performance* 1, 280-87.

––––––, **(1995)**, "The Construction of Preference," *American Psychologist*, 364-71.

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

––––––, **and** –––––– **(1983)**, "Preference Reversals: A Broader Perspective," *American Economic Review* 73, 596-605.

Starmer, Chris, and Robert Sugden (1998), "Testing Alternative explanations of Cyclical Choices," *Economica* 65, 347-61.

Sugden, Robert (forthcoming), "Reference-dependent subjective expected utility," *Journal of Economic Theory*.

Tammi, Timo (1997), *Essays on the Rationality of Experimentation in Economics, the Case of Preference Reversal*, University of Joensuu Publications in Social Sciences.

Tversky, Amos, and Richard H. Thaler (1990), "Anomalies, Preference Reversals," *Journal of Economic Perspectives*, Vol. 4, No. 2, 201-11.

Tversky, Amos, S. Sattath, and Paul Slovic (1988), "Contingent Weighting in Judgement and Choice," *Psychological Review* 95, 371-84.

Tversky, Amos, Paul Slovic, and Daniel Kahneman (1990), "The Causes of Preference Reversal," *American Economic Review* 80, 206-17.