Econometrica, Vol. 59, No. 2 (March, 1991), 425–439

OBSERVING VIOLATIONS OF TRANSITIVITY BY EXPERIMENTAL METHODS

BY GRAHAM LOOMES, CHRIS STARMER, AND ROBERT SUGDEN[1]

The preference reversal phenomenon is usually interpreted as evidence of nontransitivity of preference, but has also been explained as the result of: the difference between individuals' responses to choice and valuation problems; the devices used by experimenters to elicit valuations; and the "random lottery selection" incentive system. This paper reports an experiment designed so that none of these factors could generate systematic nontransitivities; yet systematic violations of transitivity were still found. The pattern of violation was analogous with that found in previous preference reversal experiments and is consistent with regret theory.

KEYWORDS: Decision theory, experiments, preference reversal, regret theory, transitivity, uncertainty.

1. INTRODUCTION

THE PHENOMENON OF PREFERENCE REVERSAL, first discovered by Lichtenstein and Slovic (1971) and Lindman (1971) and brought to the attention of economists by Grether and Plott (1979), presents a challenge to those who wish to explain economic behavior in terms of any theory of rational choice. Preference reversal in its classic form occurs when experimental subjects are confronted with two gambles and are asked to perform three tasks: to attach a certainty-equivalent value to each of the two gambles, and to make a straight choice between them. According to conventional theory, valuation and choice should produce the same ordering: whichever gamble is more preferred should have the higher certainty equivalent *and* be picked in the straight choice. What experiments show, however, is a systematic tendency for the ranking revealed by the valuation tasks to differ from that revealed by the choice task.

Slovic and Lichtenstein (1983) interpret this as evidence that an individual's decisions cannot be explained in terms of a single system of preferences. They suggest that preference reversals are the result of information-processing effects, and occur because the mental processes brought to bear on valuation tasks are different from those brought to bear on choice tasks. An alternative interpretation is that preferences are nontransitive: for example, Loomes and Sugden (1983) and Loomes, Starmer, and Sugden (1989) have argued that regret theory can explain preference reversals in this way. Each of these explanations involves a major departure from the received theory of rational choice.

However, three recent papers by Holt (1986), Karni and Safra (1987), and Segal (1988) have raised the possibility that the apparent violations of transitivity revealed in preference reversal experiments may result from shortcomings of

[1] The research was supported by Economic and Social Research Council Awards B 00 23 2163 and B 00 23 2236, and by the Economics Research Centre, University of East Anglia. We wish to thank the anonymous referees for their comments. Graham Loomes would also like to acknowledge the support of the Nuffield Foundation in the form of a Social Science Research Fellowship.

the experimental design. In rather different ways, these papers all argue that the experimental observations may be consistent with a theory of choice under uncertainty in which preferences are transitive, but in which other axioms of expected utility theory (EUT) are violated. In this paper we shall report an experiment which tests for systematic violations of transitivity using a design that is not vulnerable to these objections.[2]

2. THE PAPERS BY KARNI AND SAFRA, SEGAL, AND HOLT

2.1. *Karni and Safra's Argument*

Most preference reversal experiments have used a particular procedure for eliciting certainty equivalents—one first proposed by Becker, De Groot, and Marschak (1964), hereafter BDM. It works in the following way. A subject is told that he has been given a particular gamble, which he may keep and play out. Alternatively, he may try to sell the gamble back to the experimenters. He is asked to state his minimum selling price, i.e. the lowest sum of money that he is prepared to accept in exchange for the gamble. Then an 'offer' is generated by a random device. If the offer is greater than or equal to the minimum selling price, the subject is paid the full amount of the offer; otherwise he plays out the gamble. It is easy to show that a subject whose preferences satisfy the axioms of EUT will state a minimum selling price equal to the certainty equivalent of the gamble.[3] However, Karni and Safra show that if the independence axiom of EUT does not hold, then the BDM device cannot be guaranteed to elicit true certainty equivalents.

They consider a class of theories of choice under uncertainty that assume individuals' preferences over a set of lotteries to be complete, transitive, continuous, and monotonic. Their definitions have the additional effect of requiring the *reduction principle*, i.e. that compound lotteries can be reduced to simple ones by the calculus of probabilities (Karni and Safra (1987, p. 677)). Following Karni and Safra, we shall call this the class of Ω *theories*. Included among the members of this class are Machina's (1982) generalized expected utility theory, weighted utility theory (Chew and MacCrimmon (1979), and Fishburn (1983)), and expected utility theory with rank-dependent probabilities

[2] Cox and Epstein (1989) also report a preference reversal experiment which avoids the problems pointed out by Holt, Karni and Safra, and Segal. Cox and Epstein's experimental design investigates whether subjects' choices between a given pair of gambles differ according to whether the problem is framed as a single choice or as a pair of valuation tasks. Our design, in contrast, involves no valuation tasks; it investigates directly whether preferences are systematically nontransitive.

[3] To see why, consider a gamble whose certainty equivalent is v. Consider a subject choosing between the two courses of action 'state the selling price v' and 'state the selling price w', where $v > w$. Denote these two courses of action V and W, and let r be the offer generated by the random device. Then the two actions have identical consequences if $r < w$ (the subject retains the gamble) or if $r \geqslant v$ (the subject receives r). In the event that $w \leqslant r < v$, the subject receives r if he chose V but plays out the gamble if he chose W. But we know that the gamble is preferred to any value of r in the range $w \leqslant r < v$. Thus, using the independence axiom of EUT, V must be at least as preferred as W, and strictly preferred if there is a nonzero probability of an offer in the relevant range. A similar argument applies to the case in which $w > v$.

(EURDP) as proposed by Quiggin (1982) or Yaari (1987). Karni and Safra prove the following result: for every Ω theory which does not satisfy the independence axiom, it will be possible to find some pair of lotteries such that if the BDM device is used to elicit certainty equivalents and if an individual chooses his selling prices so as to achieve the outcomes he most prefers, then the ranking of the two lotteries in the individual's preference ordering will not be the same as the ranking of the minimum selling prices. This result raises the possibility that preference reversals may be caused by the incentive system used in many[4] of the relevant experiments.

What makes the preference reversal phenomenon so remarkable is that individuals are observed not merely to behave contrary to the axioms of an established theory, *but to violate that theory in a systematic and predictable direction*. In the classic experiment, the two gambles each have two outcomes. In one gamble (the $-bet), the better outcome is a relatively large sum of money, but the probability of winning it is relatively small. In the other gamble (the *P*-bet), the better outcome is a more modest sum of money, but there is a greater probability of winning it. The great majority of preference reversals take the form that the $-bet is given the higher valuation while the *P*-bet is picked in the straight choice. The opposite reversal—valuing the *P*-bet more highly but choosing the $-bet—is much less frequently observed. Although Karni and Safra do not provide a general explanation for this asymmetry, they present a numerical example in which a particular specification of the utility and probability transformation functions in EURDP, applied to a particular pair of $- and *P*-bets, generates a reversal of the commonly-observed kind.

2.2. *Segal's Argument*

Segal (1988), like Karni and Safra, argues that the BDM mechanism does not necessarily elicit true certainty equivalents if the axioms of EUT are violated. Segal, however, examines the implications of failures of the reduction principle. He suggests that two-stage lotteries are reduced to simple ones in a rather different way than is implied by the reduction principle. In Segal's theory, each second-stage lottery is replaced by its certainty equivalent. (A similar procedure is proposed by Loomes and Sugden (1986).) Segal then allows the independence axiom to be violated in the evaluation of single-stage lotteries.[5]

[4] Not all experiments have used such an incentive system. For example, in Lichtenstein and Slovic's (1971) study, Experiments I and II did not link decisions to payments. More recently, Tversky, Slovic, and Kahneman (1990) ran an experiment with two large samples, one of which answered hypothetically while the other was presented with an incentive system (not the BDM device). Substantial and systematic preference reversals were observed in all these studies, and Tversky et al. found no systematic differences between those subjects whose responses were incentive-linked and those whose responses were not.

[5] Segal adopts a somewhat unconventional definition of independence, so that the independence axiom applies only to multi-stage lotteries. On this definition, Segal's theory satisfies the independence axiom while violating the reduction principle.

Now consider an individual in a preference reversal experiment who states a selling price s for some gamble. Segal suggests that the individual may think of himself as facing the following two-stage lottery. The first stage determines whether the random offer, r, is less than s. If r is less than s, the individual plays out the gamble; otherwise, he proceeds to a second-stage lottery which determines the precise value of r. By means of a numerical example with a particular pair of \$- and P-bets, Segal shows that an individual who behaves in this way may report a higher selling price for the \$-bet even though the P-bet is preferred in a straight choice between the two.

2.3. *Holt's Argument*

In preference reversal experiments which offer real monetary incentives, it is usual to ask participants to perform a number of tasks, among which the three preference reversal tasks are included. Then one task is selected at random and the subject is paid according to his response to that task. This is the *random lottery selection procedure*. In the case of a straight choice, the subject plays out whichever gamble he chose; in the case of a valuation, the BDM procedure is used. The results of these experiments are then interpreted *as if* each subject had treated each problem in isolation. Karni and Safra's and Segal's explanations of preference reversal rest on the same 'as if' assumption.

But Holt (1986) challenges this interpretation. He assumes that subjects view the whole experiment as a two-stage lottery, where the first stage determines which task is selected and the second stage determines what they will receive when that task is played out. He then applies the reduction principle *to the whole experiment*, so that the experiment is equivalent to a choice among a large number of reduced two-stage lotteries (one for each possible combination of responses to the various tasks). Then, like Karni and Safra and Segal, he assumes that individuals have preference orderings over lotteries but does not require that these preferences satisfy the independence axiom.

On these assumptions, a subject's response to any one experimental task is not necessarily independent of the nature of the other tasks he confronts in the same experiment. Holt (pp. 512–514) argues that, if preferences are consistent with Machina's (1982) generalized expected utility theory, there may be a tendency for subjects who really prefer the \$-bet to the P-bet to choose the P-bet in the 'choice' part of a preference reversal experiment. This, he suggests, may account for the preference reversals that have been observed in experiments that have used the random lottery selection procedure.

3. REGRET THEORY AND PREFERENCE REVERSAL

Karni and Safra, Segal, and Holt have succeeded in showing that existing observations of preference reversals cannot be interpreted unquestioningly as violations of transitivity. This presents experimental researchers with a challenge. Is it possible to construct an experiment that can test for systematic intransitivities of preference that is not vulnerable to objections of the kind discussed above?

In Section 4 we shall describe an experimental design which we believe meets this challenge. The fundamental idea behind the design is that if preference reversals are a manifestation of nontransitive pairwise preferences of the kind consistent with regret theory, it should be possible to observe such violations of transitivity in an experiment consisting entirely of pairwise choice problems. In the remainder of this section we show how regret theory generates clear predictions about the direction in which pairwise choices may cycle over certain sets of three alternatives.

Regret theory applies to choices between pairs of actions (or what Savage (1954) called 'acts'). The consequence of action A_i in the event that state S_j occurs is denoted by x_{ij}, where the probability of the jth state is p_j. Consider any two actions $A_i = (x_{i1}, \dots, x_{in})$ and $A_k = (x_{k1}, \dots, x_{kn})$. Using the formulation of regret theory given in Loomes and Sugden (1987), we may define a function $\Psi(.,.)$ which assigns a real-valued index to every ordered pair of consequences. Then regret theory entails

$$(1) \qquad A_i \gtreqless A_k \quad \Leftrightarrow \quad \sum_j p_j \Psi(x_{ij}, x_{kj}) \gtreqless 0$$

where $\succ$, $\succcurlyeq$, and $\sim$ are, respectively, the relations of strict preference, weak preference, and indifference.

Three restrictions are placed on $\Psi(.,.)$:

(i) The function is *skew-symmetric* (i.e. $\Psi(x, y) = -\Psi(y, x)$ for all x, y). Notice that this implies $\Psi(x, x) = 0$ for all x.

(ii) For money consequences the function is *increasing* in its first argument.

(iii) For any money consequences x, y, z where $x > y > z$: $\Psi(x, z) > \Psi(x, y) + \Psi(y, z)$. This is the property of *regret-aversion*.[6]

We shall now show how cycles of pairwise preference can be consistent with regret theory. Consider the three actions described in Table I.

Applying (1) to pairwise choices between these actions:

$$(2) \qquad A \gtreqless B \quad \Leftrightarrow \quad p_1\Psi(a_1, b_1) + p_2\Psi(a_2, b_2) + p_3\Psi(a_3, b_3) \gtreqless 0,$$

$$(3) \qquad B \gtreqless C \quad \Leftrightarrow \quad p_1\Psi(b_1, c_1) + p_2\Psi(b_2, c_2) + p_3\Psi(b_3, c_3) \gtreqless 0,$$

$$(4) \qquad C \gtreqless A \quad \Leftrightarrow \quad p_1\Psi(c_1, a_1) + p_2\Psi(c_2, a_2) + p_3\Psi(c_3, a_3) \gtreqless 0.$$

There are two possible directions for cycles of pairwise choice: *either* B is chosen from $\{A, B\}$, C from $\{B, C\}$, and A from $\{C, A\}$, *or* A is chosen from $\{A, B\}$, B from $\{B, C\}$, and C from $\{C, A\}$. For the first type of cycle to occur, it is necessary that all three of the left-hand sides of (2), (3), and (4) be

[6] This property was called 'convexity' in Loomes and Sugden (1987). It can be shown that regret-aversion implies increasingness, but it is more convenient to treat these restrictions separately.

TABLE I

Action	Probability of state of the world p_1	p_2	p_3
A	a_1	a_2	a_3
B	b_1	b_2	b_3
C	c_1	c_2	c_3

$p_1, p_2, p_3 > 0$.

(i) $a_1 \geqslant b_1 \geqslant c_1$
(ii) $b_2 \geqslant c_2 \geqslant a_2$
(iii) $c_3 \geqslant a_3 \geqslant b_3$
with strict inequalities for one of (i), (ii), or (iii).

nonpositive, while for the second type of cycle, all three expressions must be nonnegative.[7]

Using the skew-symmetry property of $\Psi(.,.)$, the sum of those expressions is:

$$
\begin{aligned}
(5)\qquad & p_1[\Psi(a_1,b_1) + \Psi(b_1,c_1) - \Psi(a_1,c_1)] \\
& + p_2[\Psi(b_2,c_2) + \Psi(c_2,a_2) - \Psi(b_2,a_2)] \\
& + p_3[\Psi(c_3,a_3) + \Psi(a_3,b_3) - \Psi(c_3,b_3)].
\end{aligned}
$$

Because of the regret-aversion property, and because of the restrictions placed on the a, b, and c values (see Table I), the term inside each of the sets of square brackets must be nonpositive, and must be strictly negative in at least one case. Thus the whole expression in (5) is strictly negative.

In general, therefore, cycles of the type $B \succ A$, $C \succ B$, $A \succ C$ are consistent with regret theory. That is, this cycle will occur for some values of p, a, b, and c and for some $\Psi(.,.)$ function. We shall call cycles in this direction *predicted* cycles. The opposite cycle ($A \succ B$, $B \succ C$, $C \succ A$) is inconsistent with regret theory and will be referred to as the *unpredicted* cycle.

The significance of this result for the preference reversal phenomenon can be seen by looking at the three actions in Table II. This is a special case of Table I, obtained by setting $a_1 = a$, $b_1 = b_2 = b$, $c_1 = c_2 = c_3 = c$, $a_2 = a_3 = d$, and $b_3 = e$. Action C offers the payoff c with certainty. Action B has two possible payoffs, b and e; b is greater than c while e is less. Similarly, action A has two possible payoffs, a (which is greater than c) and d (which is less than c). The 'winning' outcome for action A (i.e. a) is larger than the winning outcome for B (b), but the probability of winning on B ($p_1 + p_2$) is greater than the probability of winning on A (p_1). Thus A may be thought of as representing a \$-bet and B as representing a P-bet. The structure of this set of three actions is such that every pattern of transitive preferences over the actions is consistent with regret theory

[7] We assume that, for a person who chooses, for example, A from $\{A, B\}$, it must be the case that $A \succcurlyeq B$. In other words, a person who is indifferent between A and B might choose either.

TABLE II

Action	Probability of state of the world p_1	p_2	p_3
A	a	d	d
B	b	b	e
C	c	c	c

$p_1, p_2, p_3 > 0$.
$a > b > c > d \geqslant e$.

for some $\Psi(.,.)$ function. The predicted cycle is also consistent with regret theory.[8] But, as we have shown, the unpredicted cycle is not.

Suppose someone has the predicted cycle of preferences $B \succ A$, $C \succ B$, $A \succ C$. For such a person, the certainty equivalent of A (the \$-bet) must be greater than c, while the certainty equivalent of B (the P-bet) must be less than c. Thus the \$-bet must have a higher certainty equivalent than the P-bet. But since $B \succ A$, the P-bet will be preferred in a straight choice. This is an instance of the classic form of preference reversal. In contrast, since the unpredicted cycle is inconsistent with regret theory, it is not consistent with the theory for the \$-bet to be chosen over the P-bet when the P-bet has the higher certainty equivalent.

Not all pairs of gambles in previous preference reversal experiments correspond exactly with the representation of A and B in Table II. We have constructed our gambles so that $d \geqslant e$, but there have been some cases where $d < e$, as in Pairs 1, 4, and 5 of Grether and Plott's (1979) experiments. We have also constructed our gambles so that the event in which the \$-bet gives its better consequence is a subset of the event in which the P-bet gives its better consequence. Clearly this is not the only way of juxtaposing the consequences, although it *is* in line with the design described by Grether and Plott (1979, p. 629 and Appendix) and used subsequently by Pommerehne, Schneider, and Zweifel (1982) and Reilly (1982). Most recently, Tversky, Slovic, and Kahneman (1990) have reported a large preference reversal experiment where all $\{\$, P\}$ pairs involved the same inequalities and the same juxtapositions of consequences as in Table II. Every one of these studies found a significant asymmetry between predicted and unpredicted reversals.

[8] These implications of regret theory apply to any version of Table I in which (i) A and B each offer two possible outcomes, while C offers either one or two; (ii) the winning outcome for A is greater than the winning outcome for B, which in turn is greater than the winning outcome for C (or, if C has only one outcome, the only outcome); (iii) the probability of the winning outcome for A is less than the corresponding probability for B, which in turn is less than that for C; and (iv) the losing outcomes for A and B (and for C, if C has two outcomes) are all less than any of the three winning outcomes. If these conditions hold, every possible preference ordering over A, B, C is consistent with EUT. Notice that regret theory would be identical with EUT if, instead of regret-aversion, we were to assume $\Psi(x, z) = \Psi(x, y) + \Psi(y, z)$. Because of this, it can be shown that, in addition to the predicted cycle, every possible pattern of transitive preferences over A, B, C is consistent with regret theory.

If either of the two special assumptions used in the construction of Table II is relaxed, regret theory is consistent with cycles in *either* direction (Loomes, Starmer, and Sugden (1989, p. 142)). However, since many previous experiments *do* correspond with Table II, and since the restrictions imposed in that table allow us to make an unambiguous prediction about the direction of any cycles, we confine our attention to problems with this basic structure.

4. EXPERIMENTAL DESIGN

In this section we shall first describe the design of our experiment and then show how it enables us to test for systematic nontransitivities of preference.

We began by constructing two different sets of twenty pairwise choice problems. Fifteen of these were based on five triples of actions of the general form described in Table I. (The other five questions in each set were intended to test for violations of dominance: these will be reported in a separate paper.) Each pair of actions from each triple constituted a choice problem: in terms of Table I, the problems were to choose between A and B, to choose between B and C, and to choose between C and A. Table III gives details of these triples. For ease of comparison, each triple is shown in a form that corresponds with Table I. (The numbers along the top of each matrix represent probabilities and the numbers inside the matrix represent payoffs, in UK pounds.) The triple of actions (A_1, B_1, C_1) will be called Triple 1, and so on.

Triples 1, 3, 4, 6, 8, and 9 are special cases of Table I, using different parameter values. Triples 2, 5, 7, and 10 differ very slightly from Table I: in constructing these triples we have added a fourth state of the world in which the payoff for all three actions is zero. It is easy to see that this fourth state of the world has no significance for regret theory, and that the conclusions we derived in Section 3 still apply. For each triple, probabilities and payoffs were chosen so that the predicted cycle and every possible preference ordering of A, B, C would be consistent with regret theory (see Note 8).

A total of 200 subjects took part. Subjects were allocated at random between 'Subsample I' and 'Subsample II'. The 100 subjects in Subsample I confronted one of the sets of twenty questions, which included Triples 1 to 5. The 100 in Subsample II confronted the other set of questions, which included Triples 6 to 10.

At the start of the experiment, each subject was given a booklet containing the appropriate twenty questions and some explanatory notes. The twenty problems were displayed on four sheets of paper, five problems on each sheet. On each sheet, there were three problems involving the same triple. For Subsample I, for example, the choices between A_1 and B_1, between B_1 and C_1, and between C_1 and A_1 were all displayed on the first sheet.[9] Only one triple for each subsample (Triples 5 and 10) had its three choice problems distributed

[9] The location of questions in the booklets can be worked out from the information given at the bottom of Table III. Questions 1–5 were on the first sheet, Questions 6–10 on the second, and so on.

TABLE III

PARAMETERS OF THE TRIPLES USED IN THE EXPERIMENT

Subsample I

	0.3	0.3	0.4		Expected Value
A_1	18.00	0	0		5.40
B_1	8.00	8.00	0		4.80
C_1	4.00	4.00	4.00		4.00
	0.15	0.15	0.2	0.5	E.V.
A_2	18.00	0	0	0	2.70
B_2	8.00	8.00	0	0	2.40
C_2	4.00	4.00	4.00	0	2.00
	0.4	0.3	0.3		E.V.
A_3	10.00	3.00	3.00		5.80
B_3	7.50	7.50	1.00		5.55
C_3	5.00	5.00	5.00		5.00
	0.3	0.25	0.45		E.V.
A_4	15.00	0	0		4.50
B_4	7.00	7.00	0		3.85
C_4	0	6.00	6.00		4.20
	0.18	0.15	0.27	0.4	E.V.
A_5	.00	0	0	0	2.70
B_5	7.00	7.00	0	0	2.31
C_5	0	6.00	6.00	0	2.52

$Q1$: $\{A_1, B_1\}$ $Q6$: $\{B_4, C_4\}$ $Q11$: $\{B_3, C_3\}$ $Q16$: $\{A_2, B_2\}$ $Q4$: $\{B_5, C_5\}$
$Q3$: $\{A_1, C_1\}$ $Q8$: $\{A_4, C_4\}$ $Q13$: $\{A_3, C_3\}$ $Q18$: $\{A_2, C_2\}$ $Q12$: $\{A_5, C_5\}$
$Q5$: $\{B_1, C_1\}$ $Q10$: $\{A_4, B_4\}$ $Q15$: $\{A_3, B_3\}$ $Q20$: $\{B_2, C_2\}$ $Q19$: $\{A_5, B_5\}$

Subsample II

	0.4	0.2	0.4		Expected Value
A_6	16.00	0	0		6.40
B_6	9.00	9.00	0		5.40
C_6	4.00	4.00	4.00		4.00
	0.2	0.1	0.2	0.5	E.V.
A_7	16.00	0	0	0	3.20
B_7	9.00	9.00	0	0	2.70
C_7	4.00	4.00	4.00	0	2.00
	0.3	0.4	0.3		E.V.
A_8	12.00	2.50	2.50		5.35
B_8	7.00	7.00	1.50		5.35
C_8	4.50	4.50	4.50		4.50
	0.3	0.15	0.55		E.V.
A_9	15.00	0	0		4.50
B_9	7.50	7.50	0		3.375
C_9	0	0	7.00		3.85
	0.18	0.09	0.33	0.4	E.V.
A_{10}	15.00	0	0	0	2.70
B_{10}	7.50	7.50	0	0	2.025
C_{10}	0	0	7.00	0	2.31

$Q1$: $\{B_7, C_7\}$ $Q6$: $\{B_9, C_9\}$ $Q11$: $\{B_6, C_6\}$ $Q16$: $\{B_8, C_8\}$ $Q4$: $\{A_{10}, B_{10}\}$
$Q3$: $\{A_7, B_7\}$ $Q8$: $\{A_9, C_9\}$ $Q13$: $\{A_6, B_6\}$ $Q18$: $\{A_8, C_8\}$ $Q14$: $\{B_{10}, C_{10}\}$
$Q5$: $\{A_7, C_7\}$ $Q10$: $\{A_9, B_9\}$ $Q15$: $\{A_6, C_6\}$ $Q20$: $\{A_8, B_8\}$ $Q17$: $\{A_{10}, C_{10}\}$

	1 … 30	31 … 60	61 … 100
A	8.00	8.00	0
B	18.00	0	0
	30	30	40

FIGURE 1.—Example of a choice display.

across sheets. In keeping the questions from each triple together as far as possible, our object was to reduce the scope for individuals to make nontransitive choices by mistake. We deliberately made it easy for subjects to treat the whole experiment as a single choice problem if they so wished. Our design allowed anyone who wanted to do so to look at all twenty questions before answering any, and to go back over earlier questions and revise choices before giving a set of final answers to the experimenters. In this way, subjects were given every chance to act in accordance with Holt's hypothesis.

Each subject was asked to pick a sealed envelope from a box of 100 such envelopes, and to keep the envelope sealed until the end of the experiment. Each of the envelopes in the box contained a ticket with one of the numbers 1–100; each number was in one and only one envelope, and the subjects knew this.

At the beginning of each session of the experiment, the experimenters read out the explanatory notes and made sure that subjects understood the procedure. Figure 1 reproduces a typical display from one of the booklets. (It is, in fact, Question 1—the choice between A_1 and B_1—for Subsample I.)[10] The consequences—sums of money in pounds and pence—are shown in large type, while the smaller numbers running along the top of the grid refer to the tickets in the sealed envelopes. It was explained that after making all twenty choices, each subject would roll a twenty-sided die to determine which question was to be played out for real. Suppose that this random lottery selection procedure indicated that Question 1 was to be played out by a subject in Subsample I. The experimenters would then check which action the subject had chosen on that question. At that point, the sealed envelope would be opened. A subject who had chosen A in Figure 1 would receive £8.00 if the envelope contained a ticket numbered between 1 and 60, and nothing if the ticket number was between 61 and 100. If B had been chosen, the subject would receive £18.00 if the ticket number was between 1 and 30, and nothing otherwise. The smaller numbers at

[10] In Table III, all triples are shown in a format which corresponds with Table I. However, in the experiment the actions were presented in various orders (for example, in Figure 1 the P-bet is in the upper row, with the \$-bet below it), and for some triples we put the states which resulted in the larger consequences at the right-hand ends of the displays. The letters used to identify actions to the subjects differ from those used in this paper; thus A in Figure 1 corresponds with B_1 in Table III and B in Figure 1 with A_1 in Table III. Copies of the question booklets are available from the authors on request.

the base of each column showed at a glance the chances out of a hundred that the subject's envelope would contain a ticket within that column's range.

How does this experimental design allow us to test for systematic violations of the transitivity axiom without running up against the possible objections discussed in Section 2?

We begin by assuming that the random lottery selection procedure is a valid experimental device, and thus that a subject responds to each of our twenty questions as if he or she was facing that question with certainty. Recall that the arguments of both Karni and Safra and of Segal rely on this assumption.

Now consider any triple of the form described in Table I. Suppose that the three pairwise choice problems are presented to a subject in the order $\{A, B\}$, $\{B, C\}$, $\{C, A\}$. There are eight different ways in which the subject could choose, as follows:

(i) A, B, A —consistent with $A \succ B \succ C$.
(ii) A, C, A —consistent with $A \succ C \succ B$.
(iii) B, B, A —consistent with $B \succ A \succ C$.
(iv) B, B, C —consistent with $B \succ C \succ A$.
(v) A, C, C —consistent with $C \succ A \succ B$.
(vi) B, C, C —consistent with $C \succ B \succ A$.
(vii) A, B, C —unpredicted cycle.
(viii) B, C, A —predicted cycle.

Each of the six possible orderings of A, B, and C in each of triples 1–10 is consistent with EUT. Thus each of the responses (i)–(vi) is consistent with any theory, such as generalized expected utility theory, weighted utility theory or EURDP, which includes EUT as a special case. Each of these responses is also consistent with regret theory. But if preferences are transitive, and if no errors are made, the cyclical responses (vii) and (viii) can occur only in the case of complete indifference over the three actions. There seems no reason to expect individuals who are completely indifferent to be more likely to choose (vii) than (viii) or vice versa, and so we should expect to observe these two responses equally frequently. Allowing an element of random error would not seem to change this conclusion.

By presenting our subjects only with pairwise choice problems, we have disarmed the explanation of preference reversal offered by Slovic and Lichtenstein: if the preference reversal phenomenon stems from a disparity between the way people make valuations and the way they make choices, the removal of the valuation task from the experiment should eradicate the phenomenon. For the same reason, we have disarmed the explanations offered by Karni and Safra and by Segal: our experiment does not employ the BDM device, nor any similar device for eliciting certainty equivalents. If we find that predicted cycles occur significantly more often than unpredicted ones, and if the random lottery selection procedure is valid, this must represent a systematic violation of the transitivity axiom.

So far, however, we have assumed the validity of the random lottery selection procedure. Let us now assume instead that, as Holt's hypothesis implies, a subject treats the whole twenty-question experiment as a single reducible compound lottery and then makes the most-preferred *combination* of choices. Suppose also, as Holt does, that preferences over reduced lotteries are transitive.

Given these assumptions, an individual who makes response (vii) reveals that his most-preferred compound lottery consists of actions A, B and C, each assigned a probability of $1/20$, combined with some set of seventeen other actions, which we denote by Z. But note that if this is the case, there is no reason for such an individual to choose (vii) rather than (viii) or vice versa. By the reduction principle, a compound lottery consisting of Z plus (vii) is identical with a compound lottery consisting of Z plus (viii). Thus the implication of Holt's hypothesis is that we should expect to observe the responses (vii) and (viii) with equal frequency. So this implication follows from the assumption of transitive preferences over lotteries *irrespective of whether or not the random lottery selection procedure is valid*.

That the responses (vii) and (viii) are equally likely to occur is our null hypothesis. Our alternative hypothesis is a composite one: that the random lottery selection procedure is valid, and that on each question, subjects choose according to regret theory. Since predicted cycles are consistent with regret theory while unpredicted cycles are not, we should expect (viii) responses to occur more frequently than (vii) ones.

5. RESULTS

Table IV shows the frequency of different combinations of choices for each of our ten triples of actions. It shows that the cyclical responses (vii) and (viii) together account for between 14 and 29 per cent of all observations. But they do not contribute equally: for each of the ten triples there are more predicted cycles than unpredicted ones, and in seven of these cases (Triples 1, 2, 3, 4, 5, 6, and 8) we can reject the null hypothesis that the two cycles are equally likely to occur.[11]

An aggregate view of the results can be obtained by taking the subject as the unit of observation. Any subject's response to the experiment as a whole (i.e. to the five triples he or she faced) can be assigned to one of four categories: *either* there are no cycles at all; *or* there is at least one cycle in the predicted direction, and none in the unpredicted direction; *or* there is at least one cycle in the unpredicted direction and none in the predicted direction; *or* there are cycles in both directions. The null hypothesis implies that any subject is equally likely: (a) to cycle only in the predicted direction and (b) to cycle only in the unpredicted direction. Our alternative hypothesis implies that (a) is more likely

[11] In this and subsequent tests, we use a one-tail test based on the binomial distribution. For each of Triples 1–6 and 8, $p < 0.01$; in the remaining three cases, the null hypothesis cannot be rejected at the 10 per cent level of significance.

TABLE IV

PATTERNS OF PREFERENCES OVER TRIPLES

		Subsample I: Triples 1–5					Subsample II: Triples 6–10				
		1	2	3	4	5	6	7	8	9	10
(i)	$A \succ B \succ C$	6	12	6	1	2	12	12	16	6	7
(ii)	$A \succ C \succ B$	5	9	32	8	9	3	7	10	10	7
(iii)	$B \succ A \succ C$	16	6	4	5	13	16	9	26	7	6
(iv)	$B \succ C \succ A$	9	5	1	9	10	3	3	7	12	9
(v)	$C \succ A \succ B$	4	6	10	4	6	5	8	4	13	11
(vi)	$C \succ B \succ A$	42	45	18	50	41	44	39	16	38	36
(vii)	Unpredicted	2	3	1	0	0	2	9	1	6	11
(viii)	Predicted	16	14	28	23	19	15	13	20	8	13

TABLE V

BREAKDOWN OF RESPONSES TO EXPERIMENT AS A WHOLE

Type of response	Number of subjects with this response: Subsample I	Subsample II	Total
No cycles	38	34	72
Predicted cycles only	56	42	98
Unpredicted cycles only	4	15	19
Both types of cycle	2	9	11
Total	100	100	200

than (b). Table V shows that, of 200 subjects, 98 cycled only in the predicted direction while 19 cycled only in the unpredicted direction: on this basis, the null hypothesis is decisively rejected ($p < 10^{-13}$).

6. CONCLUSION

Our experiment has revealed violations of expected utility theory that cannot be explained by random error: there is a tendency for individuals' responses to pairwise choice problems to have a cyclical pattern, and for cycles in one direction—the direction that is consistent with regret theory—to be more common than cycles in the other.

The apparent nontransitivity of preference revealed in previous preference reversal experiments has been explained in various ways—as a result of the difference between individuals' responses to choice and valuation problems, as a result of the devices used by experimenters to elicit valuations, and as a result of the 'random lottery selection' incentive system. Our experiment was designed so that none of these factors could be responsible for systematic nontransitivities. By making all questions simple pairwise choices, we have avoided the use of any mechanism for eliciting certainty equivalents. By the same means, we have eliminated the effect of any difference between 'choice' and 'valuation' responses. And the design was such that, if individuals had transitive preferences over reduced lotteries, the random lottery selection procedure could not be the cause of the asymmetric pattern of cycles we observed.

This pattern is, however, consistent with the hypothesis that the random lottery selection procedure is valid and that preferences are systematically nontransitive, as predicted by regret theory. Since we know of no other hypothesis that could explain our observations, and since there is a good deal of independent evidence which suggests that the random lottery selection procedure *is* valid,[12] our results must raise serious doubts about the descriptive validity of the transitivity axiom.

Centre for Experimental Economics, University of York, York YO1 5DD, England

School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, England

and

School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, England

Manuscript received August, 1988; final revision received February, 1990.

REFERENCES

BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1964): "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 226–232.

CHEW, S. H., AND K. R. MACCRIMMON (1979): "Alpha-nu Choice Theory: A Generalization of Expected Utility Theory," Working Paper No. 669, Faculty of Commerce and Business Administration, University of British Columbia.

COX, J. C., AND S. EPSTEIN (1989): "Preference Reversals Without the Independence Axiom," *American Economic Review*, 79, 408–426.

FISHBURN, P. C. (1983): "Transitive Measurable Utility," *Journal of Economic Theory*, 31, 393–317.

GRETHER, D. M., AND C. R. PLOTT (1979): "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 623–638.

HOLLER, M. J. (1983): "Do Economics Students Choose Rationally? A Research Note," *Social Science Information*, 22, 623–630.

HOLT, C. A. (1986): "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, 508–515.

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

KARNI, E., AND Z. SAFRA (1987): "'Preference Reversal' and the Observability of Preferences by Experimental Methods," *Econometrica*, 55, 675–685.

LICHTENSTEIN, S., AND P. SLOVIC (1971): "Reversals of Preferences Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89, 46–55.

LINDMAN, H. R. (1971): "Inconsistent Preferences Among Gambles," *Journal of Experimental Psychology*, 89, 390–397.

LOOMES, G., AND R. SUGDEN (1983): "A Rationale for Preference Reversal," *American Economic Review*, 73, 428–432.

——— (1986): "Disappointment and Dynamic Consistency in Choice Under Uncertainty," *Review of Economic Studies*, 53, 271–282.

[12] There is strong experimental evidence of what Kahneman and Tversky call the 'isolation effect' —that when faced with a two-stage problem and asked to precommit themselves to a second-stage choice to be operative if that stage is reached, most individuals make the same decision as they do when faced with the same choice without any preliminary stage (Kahneman and Tversky (1979); Tversky and Kahneman (1981); Holler (1983)).

——— (1987): "Some Implications of a More General Form of Regret Theory," *Journal of Economic Theory*, 41, 270–287.

LOOMES, G., C. STARMER, AND R. SUGDEN (1989): "Preference Reversal: Information Processing Effect or Rational Nontransitive Choice?" *Economic Journal (Conference Supplement)*, 99, 140–151.

MACHINA, M. J. (1982): "'Expected Utility' Analysis Without the Independence Axiom," *Econometrica*, 50, 277–323.

POMMEREHNE, W. W., F. SCHNEIDER, AND P. ZWEIFEL (1982): "Economic Theory of Choice and the Preference Reversal Phenomenon: A Re-examination," *American Economic Review*, 72, 569–574.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

REILLY, R. J. (1982): "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design," *American Economic Review*, 72, 576–584.

SAVAGE, L. J. (1954): *The Foundations of Statistics*. New York: Wiley.

SEGAL, U. (1988): "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review*, 78, 233–236.

SLOVIC, P., AND S. LICHTENSTEIN (1983): "Preference Reversals: A Broader Perspective," *American Economic Review*, 73, 596–605.

TVERSKY, A., AND D. KAHNEMAN (1981): "The Framing of Decisions and the Psychology of Choice," *Science*, 211, 453–458.

TVERSKY, A., P. SLOVIC, AND D. KAHNEMAN (1990): "The Causes of Preference Reversal," *American Economic Review*, 80, 204–217.

YAARI, M. E. (1987): "The Dual Theory of Choice Under Risk," *Econometrica*, 55, 95–115.