The Economic Journal, **107** (*September*), 1431–1444. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

AN EXPERIMENT ON THE PURE THEORY OF CONSUMER'S BEHAVIOUR*

Reinhard Sippel

By checking whether consumer demand satisfies the axioms of revealed preference, one can test the empirical validity of the neoclassical theory of consumer behaviour. However, applying the axioms to actual consumer purchase data is difficult, if not impossible, since it poses serious problems of both a methodological and a practical nature. After commenting on the few existing empirical studies in this field, we summarise the results from an experimental approach to revealed preference theory. Data were obtained through a controlled experiment that involved real consumption of the goods chosen. We find that most subjects violated the axioms.

When Paul Samuelson formulated revealed preference theory (cf. Samuelson, 1938) he intended to provide empirically meaningful theorems on consumer behaviour. Somewhat surprisingly, rather than using the axioms as a basis for empirical tests of the validity of the theory, subsequent research focused on the normative implications of the revealed preference approach eventually showing the equivalence of the strong axiom of revealed preference and the utility maximisation hypothesis (Houthakker, 1950). It must be kept in mind, though, that the simple static neoclassical theory of consumer demand is assumed to hold only under ideal conditions. Since these conditions are hardly ever met in economic reality, little room is left for empirical investigation. One of the major problems is that the consistency requirements imposed on the consumer by the revealed preference axioms refer to changes in demand resulting from purely hypothetical changes in prices or income. The consumer's actual behaviour over time is not restricted by the axioms unless a change of taste can be ruled out. The few studies that have been performed using real consumption data suffer from this as well as from other difficulties, and their results are far from clear-cut, so that, up to now, '(t)here is little empirical evidence concerning whether individual demands satisfy the revealed preference axioms' (Härdle *et al.* 1991, p. 1529).

Experimental economics can provide a means for a more thorough analysis of the empirical side of the revealed preference approach. Conditions very close to the theoretical requirements can be created, and the data are collected under controlled conditions. However, so far no real laboratory experiment on this topic involving human subjects has been attempted.[1] This is all the more

* I would like to thank Jim Cox, Mike Jerison, Abdolkarim Sadrieh, Reinhard Selten, seminar participants at Albany, Bonn, Bielefeld, Florence, Lausanne, Urbana-Champaign and the 7th World Congress of the Econometric Society as well as two anonymous referees and the former Managing Editor for helpful comments and criticism on earlier drafts. A revision was undertaken at the European University Institute, Florence, and benefited from its pleasant research atmosphere. Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

[1] See, however, Kagel *et al.* (1995) and the references given there for experiments using rats and pigeons as consumers. Changes in the animals' demand for food pellets and liquids after the imposition of considerable price changes are analysed using the revealed preference methodology. While there can be no

surprising since, as Pollak (1990, p. 150) points out, '(a)n "experimental" interpretation of revealed preference – one in which the economist–observer confronts the consumer with a price–expenditure situation and observes the resulting demand behavior – (...) is consistent with the positivist methodology' of Samuelson's seminal contributions.

The present paper tries to fill this gap. We report on a laboratory experiment where subjects were asked to choose goods they liked to consume facing different budget constraints. Subjects made the different choices almost simultaneously, knowing that they would actually receive one of the chosen bundles of goods. With regard to the asked-for decisions on real consumption items, the experiment is related to a number of studies on the determination of individuals' indifference curves (see MacCrimmon and Toda (1969) and further references given there). In the present experiment, actual consumption of the chosen goods is used as the incentive mechanism as well. The commodity space, however, goes far beyond that of the indifference curve studies. The random lottery incentive mechanism applied here to consumer choices on goods with a non-trivial value seems to be well suited for revealed preference tests and represents a definite improvement on the earlier studies.

The rest of the paper is organised as follows: in the next section, we briefly discuss the problems associated with empirical tests of revealed preference theory and briefly review the existing work on this subject. Section II presents a detailed account of the consumption experiment and discusses its main results. Section III concludes the paper.

I. REVEALED PREFERENCE THEORY AND EMPIRICAL APPLICATIONS

Samuelson (1938) and Houthakker (1950) have shown that neoclassical (utility maximising) demand satisfies certain consistency requirements. These requirements, known as the axioms of revealed preference, allow us to test the empirical validity of the theory.

We say that a chosen bundle of goods x^s is (directly) 'revealed preferred' to some other bundle x^t, written $x^s \quad x^t$, if and only if x^t is not more expensive than x^s at the prevailing prices p^s; formally: $x^s \quad x^t \Leftrightarrow p^s x^s \geqslant p^s x^t$, $x^s \neq x^t$. If the consumer always chooses the best bundle he can get[2] (as a utility maximiser necessarily does, of course), then, if $x^s \quad x^t$, he must never choose x^t when x^s is available, i.e.

$$x^s \quad x^t \Rightarrow \neg(x^t \quad x^s), \tag{1}$$

where $\neg$ denotes negation. This consistency requirement, the Weak Axiom of Revealed Preference (WARP), was first formulated by Samuelson (1938).

doubt about the high degree of experimental control in these studies, the animals often change their taste over the long observation periods. These experiments, though interesting in their own way, will not be discussed further in this paper.

[2] Without this additional assumption, a violation of (1) could not be called 'inconsistent'. As Sen (1993) has convincingly argued, choice behaviour is consistent, or not, only with respect to some external objective or motivation. If someone has 'a desire to violate, deliberately, the standard conditions of consistent behavior to confuse the observer (or to perplex some decision theorists)' (Sen 1993, p. 502), he will be consistent in violating (1). In a well-designed experiment, however, such a motive will never arise.

Houthakker (1950) showed that a slightly stronger version of (1) that rules out revealed preference cycles of arbitrary length is necessary and sufficient for the demand to be generated by a utility maximiser:

$$x^s \quad x^t, \quad x^t \quad x^r, \quad \ldots, \quad x^i \quad x^k \quad \Rightarrow \quad \neg(x^k \quad x^s). \tag{2}$$

Axiom (2) is known as the Strong Axiom of Revealed Preference (SARP). If a finite set of demand data does not violate (2), we can find a strictly concave utility function that generates these data.

This 'nonparametric approach to demand analysis', to quote the title of Varian's (1982) influential paper, constitutes an alternative to the more traditional approach in consumption analysis, i.e. estimating parameters for specific functional forms. Apart from formulating GARP, a slightly weaker version of SARP which allows multi-valued demand functions,[3] Varian (1982) presents algorithms that quickly determine whether consumption data are consistent with the axioms of revealed preference.

However, before applying these algorithms to real demand data, some caution is in order. Since standard demand theory is essentially static, it is implicitly assumed that the consumer acts according to an unchanged scale of preferences when he chooses the different bundles of goods. This assumption is of course most questionable if time-series of consumption data covering months or even years are analysed, as, in fact, is often the case. Any 'inconsistency' discovered might then as well be attributed to a change in taste, so that no definite conclusion can be drawn. Even if a change of preferences can be ruled out, which might be the case when observing demand behaviour over a fairly short period of time, 'inconsistencies' are likely to occur with the existence of durable goods and the possibility of storing or deep-freezing, since purchases of these goods are made only infrequently. Excluding these goods from the analysis amounts to assuming (weak) separability.[4]

While these difficulties might lead to erroneously rejecting the utility maximisation hypothesis, the opposite case, i.e. accepting the null hypothesis when it is not true, is also quite likely when real consumption data are analysed. The power of the test is very low if there is only a small number of budget set intersections. For the test to have sufficient power, it is necessary that relative prices vary greatly while income remains more or less fixed. Unfortunately, this is typically not true for real data, since relative prices vary only slightly over time, but real income increases from year to year.[5] A major advantage of an experimental analysis is that prices and income are under the control of the experimenter, so that the power of the revealed preference test can be made very high.

Despite the serious difficulties mentioned above, there is a limited number of

[3] GARP, the Generalised Axiom of Revealed Preference, is shown by Varian (1982) to be equivalent to the maximisation of a piecewise linear utility function:

$$x^s \quad x^t, \quad x^t \quad x^r, \quad \ldots, \quad x^i \quad x^k \quad \Rightarrow \quad p^k x^s \geqslant p^k x^k.$$

[4] Patterson (1991), in a non-parametric analysis of UK data, rejects the weak separability assumption for different definitions of 'durable'.

[5] See Varian (1982), p. 965, and Manser and McDonald (1988), p. 917.

empirical studies on consumer behaviour using the axioms of revealed preference. One straightforward approach is to analyse real consumption data recorded in various expenditure surveys. Households participating in these surveys write down their weekly purchases of consumption goods. Koo (1963), Mossin (1972) and Mattei (1994) are examples for this approach. In these studies most consumers have been found to behave inconsistently. However, as pointed out above, from a theoretical point of view there is no reason to expect consistency in time-series of consumption data (excluding the trivial case of non-intersecting budget planes). In addition, one should note that households do not always correctly report their purchases but tend toward underreporting, and that information regarding the price vector faced by the consumers is often incomplete. Therefore, errors in the data could cause us to reject falsely the utility maximisation hypothesis.

The need for controlled conditions when obtaining data for revealed preference tests was clearly seen by Battalio *et al.* (1973) in their study of consumer demand in a token economy set up in a psychiatric hospital. Analysis of patients' responses to large, systematic price changes from week to week showed half of them behaving inconsistently. This, however, cannot be regarded as a rejection of the theory. As Battalio *et al.* (1973) point out, there are errors in the data that might explain the inconsistencies for all but two of the subjects.[6] In addition, the experimental set-up of the token economy in a hospital did not allow a sufficient control of the relevant parameters. The token store was not the only source of consumption goods. Several patients could leave the hospital and buy goods with real currency while others received foodstuffs from visiting relatives. Moreover, for therapeutic reasons, token income or expenditure were not controlled, a fact that could easily lead to non-intersecting expenditure planes, resulting in a low power of the test for some subjects.

Obviously, the studies mentioned so far do not satisfactorily answer the question of whether consumers satisfy the axioms of revealed preference. In order to avoid the methodological difficulties of the earlier studies, we carried out two laboratory experiments, which are documented in the following section.

II. THE EXPERIMENT

In this section, we first describe the experimental set-up and then present and discuss the main results. Two slightly different experiments, which will be referred to as Exp1 and Exp2, were run. Unless otherwise indicated, the following applies to both experiments.

II.A. *The Experimental Set-up*

To obtain conditions close to the theoretical requirements and thus to provide a serious test of the theory, the consumption experiment was designed as follows.

[6] See also Cox (1997) for a more detailed analysis of the data and the effect of measurement error.

Since we wanted to analyse individual choice behaviour, only one subject at a time came to the laboratory. There, he or she had to spend a fixed amount of time (one hour) during which nothing else was allowed but the consumption of goods previously purchased from the experimenter. The subject's main task was to choose goods he or she liked to consume in the laboratory as an alternative to just sitting around and doing nothing. The prices of the goods, as well as the available budget, were denoted in an artificial unit of account and had no resemblance to actual DM-prices found in the real world. In fact, subjects did not actually pay for the goods but were asked to choose, according to their preferences, a bundle of goods they could afford given their budget and the prices. This, of course, is exactly the problem the neoclassical consumer faces when having to pick his best bundle out of the budget set.

Clearly, we need more than one demand vector per subject to address the question of consistency. One possibility would be to ask the subjects to come to the laboratory at regular intervals, e.g. every day or every week, and each time let them choose from a different budget set (with relative prices having changed). But this procedure would not allow a sufficient control of the subjects' preferences, which, as pointed out above, can be expected to change over time. Therefore, we decided to confront the subjects with 10 different budget situations almost simultaneously. Subjects were asked to state their demand in each of the 10 situations knowing that every situation had the same (one in ten) chance of being selected, with the subjects actually receiving the goods chosen in this situation. While this random lottery incentive mechanism is a standard procedure in experiments on expected utility theory, it has never before been used in experiments on revealed preference theory. This is quite remarkable since the revealed preference axioms do not relate to actual consumer choices over time but to hypothetical choices 'at the same time' (Samuelson, 1938, p. 7). The random lottery mechanism puts the subject in the required position of a consumer who has to make several choices almost simultaneously. He is repeatedly being asked, 'What would you choose if prices and income were like this?' The fact that one of the chosen bundles will actually be received gives economic significance to these otherwise purely hypothetical questions.[7]

Eight different goods were offered, listed in Table 1. They can be divided into two groups, one consisting of various food and drink items, and the other containing goods especially suited for passing the time.[8] The goods were selected in order to allow as fine a division as possible. Another consideration was that every subject should find at least some of the goods desirable. Therefore, the goods offered covered a wide range of tastes.

In both experiments, prices were chosen such that there was a large number

[7] This should induce the subject to state his true preferences. Saliency of each of the 10 decisions cannot be taken for granted but, in view of the real consequences, seems to be satisfied here. See below for remarks on the subjects' actual motivation during the experiment.

[8] One referee pointed out that there was a ninth good, 'staring at the wall'. Neglecting the demand for this good might again amount to implicitly assuming separability. But note that its price was zero in every situation, so that including it in the analysis would not change the revealed preference inequalities, and, hence, have no impact at all on our results.

Table 1
Goods Offered in the Experiment

Goods	Description	Range*
Videoclips	Watching videoclips with rock and pop music	30–60 min
Computer game	Playing 'Super Blast' (in Exp1) or 'Pinball' (in Exp2)	27·5–60 min
Magazines	Reading a selection of German newspapers and magazines	30–60 min
Coca-Cola	Cold soft drink	400–2000 g
Orange juice	Cold drink	750–2000 g
Coffee	Prepared when demanded	600–2000 g
Haribo	Popular German brand of candy, licorice, etc.	400–2000 g
Snacks	Pretzels, peanuts, etc.	600–2000 g

* In Exp1: Amount of the good available when the entire income was spent on this good only, in the situation where it was most expensive and cheapest, respectively. In Exp2, these amounts differed from subject to subject but were of the same order.

of budget set intersections giving the revealed preference test a high power. Comparing two budget situations, typically four goods were cheaper in one situation while the other four goods were cheaper in the other situation. However, budget sets were different in Exp1 and Exp2. In Exp1, two of the 10 situations (1 and 7) were virtually identical, the only difference being that prices and income were 15% higher in situation 7 than they were in situation 1. This was intended as a test of homogeneity of degree zero (implied by WARP). In order not to make the identity of situations 1 and 7 too obvious, the nominal budget available was different not only in these but in each of the 10 situations in Exp1, varying from 1,500 to 3,300. Exp2 did not test homogeneity of demand again but implemented a Slutsky-type compensation for the price changes. After having chosen their preferred bundle of goods x^0 in the first situation (denoted *S*0 and also referred to as the 'reference situation' in the instructions), subjects were compensated for the price changes in the remaining 9 situations by corresponding changes in income so that they always could afford x^0 again. Since the budget was hence endogenous in all situations but the first, this raised the possibility of strategic considerations when deciding on x^0. This problem was dealt with in two ways: first, only half of the subjects were actually informed of the compensation procedure (treatment 2b), while, to the others (treatment 2a) the budgets appeared to be exogenous. Secondly, prices for the two groups of goods were the same in *S*0 so that there was no 'cheapest' good a subject might want to choose in order to expand the other 9 budget sets.[9]

In both experiments, relative prices changed considerably between the situations though to a somewhat lesser degree in Exp2. There the compensation

[9] Specifically, the budget in *S*0 allowed a maximum consumption of 50 minutes of any of the time-consuming goods or 1,000 grams of any of the food and drink items. As it turned out, the minimum number of goods chosen in *S*0 was 2, the median being between 3 and 4.

mechanism would have translated extreme price changes into equally extreme changes in the available budget if a subject had chosen only a single good in *S*o. Even though income was not under complete control *ex ante* in Exp2, the provisions taken were sufficient to make the available budgets quite tight, forcing the subjects in both experiments to consider their choices carefully.

The experiments were divided into two parts. In the first part, after receiving verbal instructions, the subject had to fill out 10 order sheets, one for each budget situation, with prices and income varying considerably between the situations. The subject had to state his demand for each of the 8 goods offered, the only restriction being the budget restriction. Since this task required a considerable amount of calculation (multiplying prices and demand for each good and adding up to check whether the budget was exhausted), a personal computer was used for this purpose.[10] The subject simply had to enter the amounts desired, and the software informed him on the cost of the bundle and checked whether it was inside the budget set. If not, a warning message showed up on the screen. In order to help the subject not to waste anything of his budget, additional information appeared on the screen showing, for each of the 8 goods, precisely the amount that could be ordered to exhaust the budget without overrun, given the demand stated for the other 7 goods.[11] Since this information was meant only for fine-tuning the demand, it only appeared when the demand vector stated so far was sufficiently close to the budget hyperplane, namely when the corresponding expenditure was within a range of 95–105 % of the budget. This information clearly helped the subjects to avoid an annoying trial-and-error mechanism.

No time limit was imposed on the first part of the experiment. Subjects could spend as much time as they liked on their decisions and were free to compare, reconsider, and correct choices already made. This applies to all situations except *S*o in Exp2 which could not be corrected once it had been printed out.[12] When they felt that the 10 order sheets they had filled out represented their actual preferences, they asked the experimenter to start the second part. One of the 10 situations was then drawn at random using a bingo cage. The second part always lasted exactly 60 minutes, independent of how much time the subjects had spent on the first part. They received the goods chosen in the situation drawn and were allowed to consume them in any order they liked (simultaneous consumption of, say, videos and Coke was possible, too). After these 60 minutes the experiment was over, and the subjects received a fixed show-up fee of DM 25.

[10] While in Exp1 subjects had to transfer their choices from the screen to sheets of paper, in Exp2 the order sheet for each situation was printed out immediately after the decision.

[11] Subjects were not forced to spend their entire budget but were free to spend less. However, since any amount left over was not refunded, subjects clearly had an incentive to exhaust their budget (assuming non-satiation, of course).

[12] The special nature of the 'reference situation' *S*o was emphasised in the instructions to the subjects in both treatments 2a and 2b, so that no one later complained about not being able to change x^0.

Table 2

Violations of SARP/GARP in Exp1 and Exp2

	Consistent subjects (%)	Inconsistent subjects (%)	With ... violations						
			1–2	3–4	5–6	7–8	9–10	11–20	> 20
Exp1									
SARP	1 (8·3)	11 (91·7)	7	3	—	—	—	—	1
GARP	7 (58·3)	5 (41·7)	3	1	—	—	—	1	—
Exp2									
SARP	8 (26·7)	22 (73·3)	7	4	—	1	4	3	3
GARP	11 (36·7)	19 (63·3)	8	1	2	3	1	1	3

II.B. *Experimental Results*

Both experiments took place at the Laboratory for Experimental Economics at Bonn University. Exp1 involved 12 subjects, while 30 subjects participated in Exp2. The subjects were predominantly law or economics students asking for participation after a public announcement of the experiment. None of them participated in both experiments.

Table 2 summarises the experimental results.[13] In Exp1, 11 out of the 12 subjects violated SARP. Thus, only one participant can be viewed as a utility maximiser. All of the inconsistent subjects chose different bundles of goods in the identical situations 1 and 7, i.e. they showed an inhomogeneous demand and, hence, violated WARP and SARP. However, it might be argued that these subjects were indifferent to the selection between the two bundles so that choosing x^1 in situation 1 and $x^7 \neq x^1$ in situation 7 should not be called inconsistent. This is the argument underlying Varian's GARP which does not ask for a unique demand vector in every situation. Since non-homogeneity of demand was the only 'inconsistency' for 6 subjects, it leaves us with 5 subjects (41·7%) who violated revealed preference even in its weak form of GARP.[14] Although the rather strong requirement of demand homogeneity was no longer tested in Exp2, the percentage of inconsistent subjects was not reduced: in Exp2, 22 (73·3%) violated SARP,[15] and 19 (63·3%) violated GARP.[16]

While these results are not too favourable to the neoclassical theory of consumer behaviour, they deserve some closer inspection. The number of inconsistencies for each subject is typically quite low given the high power of

[13] We do not, in the following, present the results for WARP since they are very similar to the SARP results. In fact, each subject who violated SARP also violated WARP.

[14] Whether the subjects really were indifferent to the choice between the different bundles x^1 and x^7 is an open question. They were not asked explicitly about their preferences regarding the two bundles since this might have revealed to them the identity of the two situations, a fact which had to be kept unknown until all sessions of Exp1 were completed.

[15] Fourteen of these subjects actually exhibited an upward-sloping compensated demand between *So* and one or more of the other situations. Of course, the Slutsky compensation made tests of the revealed preference axioms involving the reference bundle x^0 *ex ante* more demanding than other tests. Indeed, 66 SARP violations involve *So*, significantly more than a fifth of the total number of 217 counted in Exp2.

[16] There was no significant difference in behaviour between subjects in treatments 2a and 2b. Of the 15 participants informed of the compensation, 8 violated GARP, as did 11 of the 15 others who were not informed.

the test. Since we performed $\binom{10}{2} = 45$ pairwise comparisons for each subject, and in each comparison budget sets intersected, 45 was the maximum number of SARP violations possible for a subject behaving extremely inconsistently. Yet, the median number of SARP violations in Exp2 was only 2, and for GARP the median was only 1.[17] A closer look at the actual demand data corroborates the view that the subjects did not choose randomly.[18] Every subject showed a marked preference for some of the goods while other goods were not chosen at all, even at low prices.[19] Some subjects' demand was quite price inelastic, whereas others substituted cheaper goods for their more expensive counterparts, e.g. Coke for orange juice, sometimes to the extent that they always switched from one to the other, depending upon which was cheaper in the particular situation. There can be no doubt that the subjects tried to select a combination of goods that came as close as possible to what they really liked to consume given the respective budget constraints. They spent a considerable amount of time on their decisions (typically 30–40 minutes) and repeatedly corrected entries on some of their order sheets when they reconsidered previous choices. For the 11 subjects who changed one or more of their entries in Exp2,[20] this never led to an increase but in most cases to a decrease in the number of violations. Subject 27 actually became consistent with GARP while his original choices would have violated GARP 3 times. In the majority of cases, the budget was spent entirely or at least up to a negligible remainder – the spending of which would have delivered only a fraction of a gram (or minute) of additional consumption. Subjects repeatedly expressed their discomfort with situations where preferred goods were quite expensive and hoped that another preferred situation would be drawn. From all this, we can safely conclude that subjects were highly motivated when making their decisions. Still, a majority of them violated the axioms of revealed preference.

II.C. *An Evaluation of the Results*

Given that subjects do violate the axioms of revealed preference, the question arises of how serious these deviations from optimising behaviour really are. It might be the case that the difference in 'utility' or satisfaction between a chosen bundle and another one revealed preferred to it is, in fact, hardly noticeable for the subject. We might then regard the inconsistency of not choosing the seemingly preferred bundle as being of minor importance. There is, however, no way to tell whether a deviation is serious or not since we do not

[17] In Exp1, the medians were even lower and both smaller than 1. Note, however, that in Exp2 there were actually 2 subjects (subjs. 5 and 7) who showed 45 violations of SARP, i.e. for these subjects we have that for any $x^s \neq x^t, s, t = 0, \ldots, 9$, both x^s is (indirectly) revealed preferred to x^t and x^t is (indirectly) revealed preferred to x^s, so that no preference pattern emerges from their behaviour at all. In particular, the substitution effect of the compensated price changes is non-negative in 7 (subj. 5) and 6 (subj. 7) out of 9 cases.

[18] These data are not reproduced here but are available from the author upon request, as are the instructions and software used in the experiment.

[19] Some subjects explicitly stated to the experimenter that they disliked certain goods, such as coffee or playing a computer game; consequently, they never chose these goods.

[20] Unfortunately, in Exp1 the original choices were not recorded.

know the 'true' preference structure. In fact, without this knowledge we cannot even tell what a deviation is and what behaviour is in accordance with these preferences.

Notwithstanding these problems, several authors have suggested measures for rating revealed preference violations. A particularly simple and widely used measure is the Afriat efficiency index (see Afriat, 1973; Varian, 1993). Acting inconsistently, i.e. choosing some consumption bundle x^t instead of another bundle x^s which has been revealed preferred to it and which is affordable as well, obviously amounts to wasting income. The smaller the ratio $p^t x^s / p^t x^t$, where $x^s \quad x^t$, the greater the inefficiency. If this ratio, called the Afriat efficiency index and denoted by e, is close to 1, the waste of income might be insignificant for the consumer. This argument suggests reformulating the revealed preference relation in a weaker form, so that $x^s \quad x^t$ if and only if x^t is 'distinctly' cheaper than the chosen bundle x^s: $x^s \quad x^t \Leftrightarrow e\, p^s x^s \geqslant p^s x^t$, $x^s \neq x^t$. If we set $e = 0{\cdot}9$, for example, then x^s has to be more than 11% more expensive than x^t before we conclude that $x^s \quad x^t$. Table 3 shows that the

Table 3

Inconsistent Subjects and Inconsistent Random Demand (Violating GARP) Depending upon the Afriat Efficiency Index e

	Number of inconsistent subjects (%)		Test power (inconsistent random demand) in %	
e	Exp1	Exp2	Exp1	Exp2
1·0	5 (41·7)	19 (63·3)	61·3	97·3
0·99	3 (25·0)	8 (26·7)	46·8	65·2
0·95	1 (8·3)	3 (10·0)	16·8	12·8
0·90	1 (8·3)	1 (3·3)	1·5	0·4

number of inconsistent subjects is substantially reduced if the relation is weakened in the above manner. With an efficiency of 95%, no more than 10% of the subjects remain inconsistent. This might suggest that the theory is actually confirmed, if we allow for a small optimising error.

Note, though, that introducing Afriat inefficiency severely reduces the power of the test. The lower e is, the fewer the preferences which are revealed.[21] Hence, the theory makes fewer (or poorer) predictions, and violations are less frequent. Simulations show that the frequency of GARP violations by purely random demand is reduced in the very same way (see the right-hand column of Table 3).[22] While these 'artificial consumers' are much more inconsistent than the real subjects are at $e = 1$, most of them satisfy GARP at $e = 0{\cdot}95$. In

[21] Note in particular that, with $e = 1 - \epsilon, \epsilon$ being an arbitrarily small positive number, it is no longer 'revealed' that $x^s \quad x^0$ for all $s = 1, \ldots, 9$ in Exp2, though this, of course, is true by virtue of the compensation mechanism for all $x^s \neq x^0$.

[22] In the simulations, 1,000 demand vectors were created from randomly determined budget shares subjected to the price vectors and income changes of Exp1 and Exp2, respectively; see Sippel (1996) for details.

other words: a non-parametric test of optimising behaviour with 95% efficiency has practically no power against the alternative of purely random choice.

These unpleasant consequences of allowing seemingly small amounts of Afriat inefficiency force us to look for a different procedure if we want to account for possible stochastic influences and errors in decision making. A failure to satisfy GARP might be the result of a small error in the consumer's choices, and there might exist demand vectors close to the actual ones that satisfy GARP.[23] We will try to account for the well-known fact of variability in subjects' choices[24] by perturbing the reported demand vectors and seeing how sensitive our results are to these perturbations.[25] If consistent subjects become wildly inconsistent after slight changes in their demand, and *vice versa*, then the significance of the results might be in question. Additionally, if we can find demand vectors satisfying GARP sufficiently similar to the actual ones violating it, this subject can be seen as being 'almost consistent'.

The sensitivity analysis proceeded as follows: for each subject in Exp2, we ran 100 simulations where demand was perturbed in situations 1 to 9 and calculated the number of GARP violations under the perturbations. The new demand vectors were determined randomly but had to satisfy three conditions: (i) they had to cost the same as the original one, (ii) components which were zero in the original vector were zero in the perturbed vector, and (iii) the angle between the two vectors did not exceed a prespecified value. These conditions assured that the new demand was feasible and in a well-defined neighbourhood of the original one. The simulations were repeated for different values of the maximum angle in order to study the effects of increasing the variability of choices. An analysis of the demand differences between the identical situations 1 and 7 of Exp1 provides information on the magnitude of variability shown by subjects in our setting.[26] The angles between the two vectors x^1 and x^7 cover the full range between 0° (for the one subject with homogeneous demand) and 90° (for a subject who demanded Magazines, Orange Juice and Haribo in one situation but Videos, Coke and Snacks in the other). The latter subject, however, is clearly an exception, since the distribution of angles is far from uniform: all but one of the angles are smaller than 32°, the mean is 19·3°, and the median is 12·7°. Hence, in general, variability in choices is not too high so that the two demand vectors are quite similar.[27]

The results of the sensitivity analysis are summarised in Table 4. It shows the average number of GARP violations for five groups of subjects in Exp2 for four different values of the maximum perturbation angle (0–38·5°). It can be seen

[23] 'Trembles' such as unintentionally entering incorrect amounts on the order sheets are but one kind of possible error in our case. More generally, we will assume that, because of some unspecified stochastic process, there is no longer a one-to-one correspondence between preferences and choices. How to reconcile this with the very notion of 'revealed preference' is of course another matter.

[24] That is, subjects fail to always replicate their choices when faced with the same decision problem again.

[25] Manser and McDonald (1988) and Patterson (1991) perform similar sensitivity analyses.

[26] Since the demand differences between these two situations might also be caused by money illusion, this measure is an upper bound on the variability to be expected between truly identical situations.

[27] For illustration, here is the median subject's demand: $x^1 = (25, 8, 0, 0, 200, 80, 0, 150)$ and $x^7 = (30, 0, 0, 0, 200, 150, 0, 200)$.

Table 4

Sensitivity Analysis: Average Number of GARP Violations for Subjects in Exp2 (grouped according to their actual number of violations) Depending on the Maximum Perturbation Angle

	Subjects with ... violations				
Angle (°)	0 ($N = 11$)	1–2 ($N = 8$)	3–6 ($N = 3$)	7–10 ($N = 4$)	> 10 ($N = 4$)
0	0	1·38	4·67	8·0	26·75
5	0·97	2·95	5·07	9·8	27·9
19·3	1·72	4·09	8·53	13·58	26·05
38·5	3·03	5·15	10·57	16·5	25·58

that the average number of violations does not change much under the perturbations.[28] Subjects with only a few (or no) inconsistencies remain on this low level after perturbations of their demand. Subjects with a higher number of violations remain on this high level or become even more inconsistent after the perturbations. In general, the number of inconsistencies increases with the size of the maximum perturbation angle. This, of course, could be expected for the consistent subjects since the newly constructed demand is ever farther away from its optimal value. But the fact that this is also true for the inconsistent subjects[29] seems to indicate that variability cannot sufficiently explain the violations. Another indication for this view is that we were able to find consistent demand vectors for only 8 out of the 19 inconsistent subjects, even though the maximum variability in the perturbations (38·5°) was about twice the mean variability of subjects in Exp1. Since these 8 subjects coincide with the subjects who violated GARP only once or twice, we might have good reason to call them 'almost consistent'. The others' behaviour, however, seems to be guided by principles other than optimising.

We conclude that the evidence for the utility maximisation hypothesis is at best mixed. While there are subjects who appear to be optimising, the majority of them do not. The high power of our test might explain why our conclusions differ from those of other studies where optimising behaviour was found to be an almost universal principle applying to humans and non-humans as well. In contrast to this, we would like to stress the diversity of individual behaviour and call the universality of the maximising principle into question.

III. CONCLUSION

The revealed preference approach to consumer theory, which originated with Samuelson's (1938) seminal paper, is based on the assumption that 'the individual guinea-pig, by his market behaviour, reveals his preference pattern

[28] In fact, this and the following also hold on the individual subject level. Lack of space does not permit a more disaggregated presentation of the results here.

[29] Except for those with a very high number of violations. Since they are already close to the maximum possible number of violations, they can hardly become more inconsistent with a different demand.

– if there is such a consistent pattern' (Samuelson, 1948, p. 243). Thus, it is formulated in a positivist vein and is, in principle, subject to empirical refutation. However, such an empirical test only makes sense in an experimental environment where tight control of the relevant parameters, in particular the consumer's preferences, can be maintained. The experiment reported here was designed to reflect the fact that revealed preference theory is concerned with hypothetical choices rather than actual choices over time. In contrast to earlier experimental studies, the possibility that the different choices are made under different preference patterns can almost be ruled out. We find a considerable number of violations of the revealed preference axioms, which contradicts the neoclassical theory of the consumer maximising utility subject to a given budget constraint. We should therefore pay closer attention to the limits of this theory as a description of how people actually behave, i.e. as a positive theory of consumer behaviour.[30] Recognising these limits, we economists should perhaps be a little more modest in our 'imperialist ambitions' of explaining non-market behaviour by economic principles.

Another conclusion concerns the usefulness of the experimental set-up for further research on properties of consumer demand. It turned out that subjects were highly motivated when making their decisions. Although their choices were not related to monetary payoffs but to the possibility of consuming a specific bundle of goods in the laboratory, the latter proved to be of a non-trivial value, inducing the subjects to weigh their alternatives carefully. Thus, an experimental analysis of consumer demand along the lines laid out here can be a fruitful approach, especially in cases where real-economy consumption data are either not reliable enough or simply not available.

Universität Bonn

Date of receipt of final typescript: December 1996

References

Afriat, S. N. (1973). 'On a system of inequalities in demand analysis: an extension of the classical method.' *International Economic Review*, vol. 14, pp. 460–72.

Battalio, R. C., Kagel, J. H., Winkler, R. C., Fisher, E. B., Basmann, R. L. and Krasner, L. (1973). 'A test of consumer demand theory using observations of individual consumer purchases.' *Western Economic Journal*, vol. 11, pp. 411–28.

Cox, J. C. (1997). 'On testing the utility hypothesis.' Economic Journal, vol. 107, pp. 1054–78.

Härdle, W., Hildenbrand, W. and Jerison, M. (1991). 'Empirical evidence on the law of demand.' *Econometrica*, vol. 59, pp. 1525–49.

Houthakker, H. S. (1950). 'Revealed preference and the utility function.' *Economica*, vol. 17, pp. 159–74.

Kagel, J. H., Battalio, R. C. and Green, L. (1995). *Economic Choice Theory*. Cambridge: Cambridge University Press.

Koo, A. Y. C. (1963). 'An empirical test of revealed preference theory.' *Econometrica*, vol. 31, pp. 646–64.

MacCrimmon, K. R. and Toda, M. (1969). 'The experimental determination of indifference curves.' *Review of Economic Studies*, vol. 36, pp. 433–51.

Manser, M. E. and McDonald, R. J. (1988). 'An analysis of substitution bias in measuring inflation, 1959–85.' *Econometrica*, vol. 56, pp. 909–30.

Mattei, A. (1994). 'La cohérence des choix des consommateurs suisses.' *Revue Suisse d'Économie Politique et de Statistique*, vol. 130, pp. 3–20. English version: 'A non-parametric test of consumer behavior using panel data.' Université de Lausanne, Deep Discussion Paper no. 9112.

[30] See Thaler (1980) for further examples of divergence between normative theory and actual behaviour of consumers.

Mossin, A. (1972). 'A mean demand function and individual demand functions confronted with the weak and the strong axioms of revealed preference: an empirical test.' *Econometrica*, vol. 40, pp. 177–92.
Patterson, K. D. (1991). 'A non-parametric analysis of personal sector decisions on consumption, liquid assets and leisure.' ECONOMIC JOURNAL, vol. 101, pp. 1103–16.
Pollak, R. A. (1990). 'Distinguished Fellow: Houthakker's contributions to economics.' *Journal of Economic Perspectives*, vol. 4, 2, pp. 141–56.
Samuelson, P. A. (1938). 'A note on the pure theory of consumer's behaviour.' *Economica*, vol. 5, pp. 61–71.
Samuelson, P. A. (1948). 'Consumption theory in terms of revealed preference.' *Economica*, vol. 15, pp. 243–53.
Sen, A. K. (1993). 'Internal consistency of choice.' *Econometrica*, vol. 61, pp. 495–521.
Sippel, R. (1996), 'A note on the power of revealed preference tests with Afriat inefficiency.' SFB 303, Universität Bonn, Discussion Paper No. A-528.
Thaler, R. (1980). 'Toward a positive theory of consumer choice.' *Journal of Economic Behavior and Organization*, vol. 1, pp. 39–60.
Varian, H. R. (1982). 'The nonparametric approach to demand analysis.' *Econometrica*, vol. 50, pp. 945–73.
Varian, H. R. (1993). 'Goodness-of-fit for revealed preference tests.' Mimeo, University of Michigan.