Coping with Preference Anomalies in Cost–Benefit Analysis: A Market-Simulation Approach

ROBERT SUGDEN
School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK
(e-mail: r.sugden@uea.ac.uk)

Abstract. This paper proposes a methodological strategy for cost–benefit analysis (CBA) which does not require the assumption that individuals' preferences satisfy standard coherence conditions, and so renders CBA immune to the problems generated by preference anomalies. The proposal treats CBA as an exercise in market simulation, based on the measurement of surplus. Anomalies occur when surplus measurements vary according to the hypothetical payment mechanism used. In such cases, the mechanism that is the "closest market analogue" should be used. This approach is used to resolve problems associated with some familiar anomalies, including inconsistencies between "citizen" and "consumer" valuations, and endowment effects.

Key words: anomalies, cost–benefit analysis, market simulation

JEL classifications: D61, D63, Q51

Economics has conventionally assumed that individuals have preferences over all economically relevant outcomes. These preferences have been assumed to be *stable* (that is, they are not liable to sharp changes over short periods of time), *context-independent* (that is, different choice problems, and different contextual "framings" of what economic theory would regard as the same problem, evoke the same underlying preferences), and *internally consistent* (that is, they satisfy such standard conditions as completeness, transitivity and the independence axiom). The assumption that preferences have these properties – the assumption of *preference coherence* – underpins much of the theory that is used to explain and predict economic behaviour. It is also a fundamental component of most normative work in economics, since such work usually takes as its objective the maximisation of some increasing function of the well-being of individuals, and treats each person's preference ordering as an ordinal measure of her well–being. When, as is common in environmental applications of cost–benefit analysis (CBA), individuals'

preferences are inferred from survey data, consistent analysis is possible only if the coherence assumptions imposed by the theory being used are reflected in those data.

For many years, preference coherence was taken for granted by economists. However, there is now a large body of evidence of *preference anomalies* – that is, of systematic deviations between actual decision-making behaviour and the implications of standard assumptions about preferences.[1] These anomalies pose severe problems for normative economics. In CBA, the problems are particularly acute when, as is often the case, anomalies show up in the survey data from which preferences are to be inferred. Particularly intractable problems have been caused by disparities between willingness-to-accept (WTA) and willingness-to-pay (WTP) valuations and by apparently insufficient sensitivity of stated valuations to the scale of the good being valued.[2]

The significance of anomalies – for economics in general and for CBA in particular – is currently the subject of vigorous debate. Some commentators interpret the evidence as revealing fundamental deficiencies in the received theory of decision-making behaviour. Others see anomalies as epiphenomena – as effects which may be replicable in certain kinds of laboratory experiments, but which result from errors and misperceptions which people learn to avoid in decisions which really matter to them, and which can be screened out by well-designed survey instruments.[3] That debate provides the background for this paper, but – I emphasise this point – the paper is *not* a contribution to it. I do not set out to argue for any particular position on the significance or otherwise of anomalies. Instead, I propose an approach to CBA which uncouples it from the assumption of preference coherence.

Of course, the reader's judgement of the value of this strategy is likely to depend on his or her assessment of the significance of anomalies. Most economists, I imagine, would be willing to consider dropping the assumption of preference coherence only as a last resort, as a drastic remedy to otherwise intractable problems. My own interest in this approach is in part a reflection of my belief that anomalies are not epiphenomena, and that economics needs a form of CBA which is robust against a wide range of deviations from the received theory of decision-making. Nevertheless, my proposal does not depend on any specific assumptions about failures of preference coherence. To the contrary, it is entirely compatible with conventional assumptions about preferences: if those assumptions hold, the proposal simply endorses current CBA methodology.

I claim that my proposal allows CBA to cope with many different ways in which the assumption of preference coherence might fail. To support this claim, I consider various patterns of behaviour which, if observed in reality, would count as anomalies and would create problems for conventional CBA; I explain how my proposal would respond to these problems. Clearly, if I am

to test the robustness of my proposal, it is particularly valuable to consider would-be anomalies which constitute major deviations from received theory. My aim is to find CBA methods which can cope with the sort of anomalies that, on the most pessimistic interpretations of the current evidence, we might expect to find. Unavoidably, therefore, I will be considering patterns of behaviour or "effects" whose status as empirical regularities is controversial. My argument does not depend on the claim that each of these effects does in fact occur. However, I take these effects to be more than mere theoretical possibilities. For each of them, there is enough evidence to give at least some credence to the hypothesis that they occur. Thus, other things being equal, we have reason to value a CBA methodology which can cope with them if they do.

In Section 1, I explain the fundamental principle which underlies my proposal: that CBA should be interpreted as an attempt to simulate the workings of ideal competitive markets in situations in which, in reality, there is market failure. On this account, CBA is concerned with the measurement of economic surplus. In Section 2, I present a definition of economic surplus which does not presuppose that individuals' preferences are coherent. I show how measurements of surplus, so defined, are possible, given only minimal assumptions about the "price sensitivity" of consumers and the about predictability of their behaviour in the aggregate. In Section 3, I argue that most of the anomalies which cause problems for CBA arise because individuals' preferences are context-dependent. Since this implies that surplus is context-dependent too, the key to resolving the problems is to find a general criterion for selecting the contexts in which measurements of surplus are made for the purposes of CBA. Guided by the principle of market simulation, I propose a "closest market analogue" criterion. The remainder of the paper looks at implications of this proposal. Section 4 considers whether, when valuations are elicited by stated-preference methods, questions should be posed in a "citizen" or "consumer" frame. Sections 5 and 6 consider how CBA should deal with cases in which preferences over given options vary according to the reference point adopted by the individual.

1. What Is Cost–Benefit Analysis?

When economists explain the principles of CBA, three alternative rhetorical strategies are commonly used. The *welfarist* strategy is to represent the object of the exercise as the maximisation of social welfare. On this account, an individual's preferences are a source of information about her well-being; CBA uses preference data as a means of reaching conclusions about social welfare. The *public choice* strategy sees the purpose of CBA as to identify decisions that would emerge from an idealised political process. When CBA

is explained in these terms, the failure of conventional methods of voting to take account of strength of preference is usually treated as a departure from the ideal. So too is the phenomenon of strategic voting.[4] The abstract ideal is a collective choice mechanism which aggregates individuals' sincere preferences, taking account of their relative strength. If money valuations are accepted as measures of strength of preference, CBA can be interpreted as simulating such an ideal mechanism. The *market simulation* strategy sees the purpose of CBA as to simulate the workings of an ideal competitive market. On this account, public goods, externalities and other forms of market failure are treated as departures from the ideal. (Throughout this paper, I use the term "ideal competitive market" to refer to the classic model of a perfectly competitive economy in which there is no market failure.) By virtue of the first fundamental theorem of welfare economics, the equilibrium state of an ideal competitive market has the property of Pareto-efficiency. CBA is interpreted as a method of generating the kind of outcomes that are generated by ideal competitive markets – that is, outcomes that are Pareto-efficient – but which can be applied in situations of market failure.

If each individual has just one preference ordering, governing her behaviour both in markets and in the political process, and if suitable definitions of "ideal" political processes and "ideal" competitive markets are used, these three rhetorical strategies are mutually consistent. From the welfarist perspective, an individual's preferences can be interpreted as measures of her welfare, while from the public choice perspective they can be interpreted as the judgements she makes as a citizen. On either interpretation, CBA arrives at recommendations by aggregating preferences across individuals. On conventional definitions of social welfare, Pareto-efficiency is a necessary condition for the maximisation of social welfare. In this sense, market simulation is consistent with welfarism.

But what if individuals do *not* have coherent preferences, either because their choices reveal systematic preference anomalies, or because the judgements they wish to register in the political process do not coincide with the preferences that govern their choices as consumers? Then these three ways of understanding CBA diverge. For someone who favours the welfarist interpretation of CBA, a natural response is to look for new measures of well-being which do not tie well-being as tightly to preference as welfare economics has traditionally done. A proponent of the public choice interpretation might try to enrich conventional methods of preference elicitation so that they take on more of the character of collective deliberation, perhaps on the model of citizen juries. In this paper, however, I explore the implications of maintaining the principle of market simulation. I propose a form of CBA which, even if preferences are not coherent, can still be interpreted as a simulation of markets.

The reader may ask: Why simulate markets, unless it can be shown that markets generate well-being? What is the point of a measure of costs and benefits which does not tell us how to increase social welfare, and which is not a method of decision-making?

My answer is that CBA as market simulation is ultimately a set of accounting conventions. These conventions allow information about policy options to be organised in a standardised and systematic framework. The information so organised is likely to be *useful* to anyone who is deliberating about how well the various options measure up against her own system of values, or who is taking part in the process of choosing between those options. In order for CBA to be useful in this sense, it does not have to tell us which option is welfare-maximising according to some particular (and inevitably contested) definition of social welfare. Nor does it have to pre-scribe which option should be chosen. What is necessary is that the infor-mation generated by CBA is *meaningful* and *relevant*: it should be information about some general feature of policy options, and that feature should be one which, when viewed from any of a range of different systems of values, can be recognised as relevant for decision-making. In Section 2, I shall argue that CBA as market simulation satisfies these criteria, by virtue of providing information about economic surplus.

However, in appraising any set of accounting conventions, we should be prepared to accept some degree of arbitrariness. Independent of the relative merits of different CBA methodologies, there is some value in standardisa-tion. If the same accounting conventions are used in many different cases, the statistics they generate tend to become more informative, merely by virtue of the comparisons that can be made across those cases. For example, most economists would agree that the conventions used to measure national income are, in significant respects, arbitrary. Nevertheless, national income data are informative. Our ability to interpret these data in any particular case is enhanced by the large body of national income data that now exists for different countries and different time periods. For the same reasons, our ability to interpret the results of any particular application of CBA is enhanced by our being able to make comparisons with other applications of the same methodology.

Standardisation is particularly important if CBA is to be used within a decision-making process that involves the interaction of parties with con-flicting interests and values. The results of CBA studies are often used, not only to guide the process of policy making, but also to justify particular policy choices to the general public. Public sector decisions often involve more than one level of government (local, regional, national or suprana-tional); increasingly, they also involve partnerships between government and private firms. In such decision-making environments, CBA results are used in negotiations between parties. When one tier of the government seeks

co-funding for a project from another tier, or when a private firm seeks co-funding from government, CBA may be used to support a claim that that project is worth supporting. If CBA is to be used in public debate and in negotiations, openness and credibility are essential. The existence of a generally accepted set of accounting conventions makes it easier for a CBA that has been sponsored by one agency or interest group to be scrutinised by another. Further, the credibility of the results of one study can, to some degree, be tested by comparing them with those of similar studies which have used the same methodology.

In proposing that CBA is interpreted as market simulation, I am proposing that economists coordinate on a particular set of accounting conventions for measuring costs and benefits. I am conscious that any such proposal can be read as "Be reasonable – do it my way". But I hope to persuade the reader that my proposal is not idiosyncratic. As a foundation for CBA, market simulation is salient in two significant ways. First, it follows a precedent: as I have argued, existing CBA methodology can be, and has been, interpreted as market simulation. My proposal maintains that interpretation. Second, in political debate in most developed countries there is now a widespread presumption in favour of competitive markets as the preferred mechanism for allocating resources in relation to consumer goods. The idea of market simulation – of basing CBA on the same concepts of cost and benefit that govern the allocation of resources in competitive markets – is in harmony with that presumption.

2. Surplus

On the interpretation I propose, CBA measures *surplus*. Surplus, as I define it, is a property of transactions in which goods are transferred between individuals. An individual gains surplus from a transaction to the extent that the goods she receives more than compensate her for the goods she gives up. In order to measure surplus, we have to choose some good to use as the standard of value. Which good we choose is to some extent a matter of convention. However, we need it to be a good such that we can be confident that, whatever their other preferences may be, people always prefer to have more rather than less of it; and we need it to be finely divisible. Money is the obvious choice: from now on I shall use money as the standard of value.

If an individual has coherent preferences, the concept of compensation in the definition of surplus can be interpreted in terms of indifference surfaces: to compensate an individual for some loss is to move her back to her original indifference surface. In this case, surplus corresponds with the welfare-economic concept of a *compensating variation*.[5] But it is not necessary to assume

preference coherence in order to define surplus. A much weaker assumption of *price sensitivity* is sufficient.

Consider any given specific would-be transaction T, which some particular person at some particular moment is considering whether to make. We can define a family of hypothetical transactions, each of which is the effect of combining T with an additional net outlay (positive, zero, or negative) of money. Let (T, x) denote the transaction which combines T with a net outlay of x. The assumption of price sensitivity is that if, *at that moment*, that person would be willing to make some transaction (T, x'), rather than not transact at all, then she would be willing to make instead any other transaction (T, x''), where $x'' < x'$. Now let x^* be the highest value of x such that the individual is willing to make the transaction (T, x). Then x^* is the surplus that the person earns on the transaction T. Notice that, because surplus is measured in units of money, it can meaningfully be added across individuals. In particular, we can talk about the total surplus that two or more trading partners earn on a given transaction.

An individual is price-sensitive if, in buying any given good at any given moment, she prefers to pay less money rather than more, and if in selling any given good at any given moment, she prefers to be paid more money rather than less. Notice that this assumption does *not* imply that the individual's preferences are coherent. For example, consider some individual who is endowed with one unit of some good, and who at 10:30 on a particular day is offered the opportunity to sell it. Suppose she would sell it at any price greater than or equal to £10, but not at any lower price. This pattern of behaviour is price-sensitive. Suppose that the price is in fact £10, and she sells. At 10:35 she is offered the opportunity to buy the good back. Suppose she would now buy it at any price less than or equal to £5, but not at any higher price. This behaviour, too, is price-sensitive. However, if we assume that more money is preferred to less, the individual's behaviour as a whole does not satisfy the assumption of preference coherence. (If a gain of £10 just compensates her for giving up the good, then, having given up the good and received the £10 in exchange, getting the good back should just compensate her for paying back the £10. So her WTP for the good at 10:35 should be £10, not £5.) It seems that, if we are to rationalise this behaviour in terms of references, we must infer *either* that the individual's preferences are unstable (there is a sharp change between 10:30 and 10:35) *or* that they are context-dependent (the preferences revealed in the buying problem are different from those revealed in the selling problem) *or* that they violate standard consistency conditions. In other words: this is a form of price-sensitive behaviour that cannot be rationalised in terms of the received model of coherent preferences.

This is not a fanciful example. Many of the standard methods of eliciting WTP and WTA valuations work by offering each respondent a series of

dichotomous choices involving different buying prices (in the case of WTP) or selling prices (in the case of WTA) for a given benefit, thus indirectly revealing the highest buying price the respondent is willing to pay or the lowest selling price he is willing to accept. It is only because respondents can be relied on to be price-sensitive that these methods elicit unambiguous WTP and WTA valuations. Yet experiments and surveys using such methods have repeatedly found systematic disparities between the two forms of valuation. In addition, there is strong experimental evidence that individual decision-making behaviour is subject to stochastic variation. If a given individual faces exactly the same choice task twice within a single experiment, with perhaps only a few minutes elapsing between the tasks, there is a predictable degree of random variation between the two responses; for a given dichotomous choice between lotteries, the proportion of subjects which reveals different preferences on the two occasions is typically of the order of 25 per cent (Camerer 1989; Starmer and Sugden 1989; Hey and Orme 1994, p. 1269). It seems that, at the level of the individual, apparently random reversals of preference are a fact of life.

Conventional welfare economics analyses the capacity of ideal competitive markets to create surplus, given the assumption of preference coherence. Two theoretical results are of particular significance. First, the first fundamental theorem of welfare economics tells us that competitive equilibrium is Pareto-efficient. This result implies that, starting from a competitive equilibrium, no further transactions exist which would create non-negative surplus for all individuals and strictly positive surplus for some.

The second result concerns the comparative statics of competitive equilibrium. In an ideal competitive market, every marginal transaction (that is, every transaction which induces only infinitesimal changes in market prices) which creates an increase in surplus for each trading partner involved in it also generates a net increase in surplus for all economic actors taken together. It is important to understand exactly what this result means. It does *not* imply that market transactions have no effect on individuals who are not party to them. To the contrary, transactions generally *do* have positive and negative effects on the surplus of third parties. What we know is that these effects cancel out in the aggregate. When a transaction takes place, surplus is *transferred* between third parties, and the gross value of this transfer is of the same order of magnitude as the value of the original transaction.[6]

These two results can be interpreted as showing, in different ways, that competitive markets realise opportunities for the creation of net surplus. The first result shows that, in competitive equilibrium, no such opportunities remain unrealised. The second result shows that, whenever trading partners in a competitive market realise gains from trade amongst themselves, they thereby create a net increase in surplus for all economic actors taken toge-

ther. These results are formal counterparts of the familiar informal claim that markets are effective mechanisms for *wealth creation.*

In deriving these conclusions, conventional welfare economics defines surplus in terms of preferences, using the concept of compensating variation, and assumes that preferences are coherent. But, I have argued that surplus can be defined without using that assumption. One might then ask whether the results of conventional welfare economics still apply: do competitive markets in private goods realise opportunities for the creation of net surplus, even if individuals' preferences are not coherent? In another paper (Sugden 2004), I argue that the answer is "Yes".

The argument is conducted in terms of a stylised model of a competitive exchange economy. In this economy, there are two kinds of actors: *consumers* and *traders*. Consumers come to the market with endowments of various private goods, one of which ("money") serves as the medium of exchange. Traders make public offers to buy and sell goods at prices denominated in units of money. By selling and buying goods at these prices, prompted by their current desires, consumers adjust their holdings of goods. Those desires need not be capable of being represented as coherent preferences: all that is assumed about consumers' behaviour is that, in the aggregate, it is predictable by traders, and that, at every moment, consumers are price-sensitive. Traders seek to make profits, measured in money, by buying at low prices and selling at high prices. The market is defined to be in a state of *free-entry equilibrium* if no active trader makes a loss, and if no potential trader could make a profit by entering the market. The main result is this: in a free-entry equilibrium, provided there is at least some trade in each non-money good, there is for each such good a single price, at which all consumers are free to buy and sell whatever quantities they choose, and at which total purchases by consumers are equal to total sales by consumers. Thus, a market-clearing price vector – the essential characteristic of competitive equilibrium, as conventionally defined – is induced by competition among profit-seeking traders. It follows from this result that, in a free-entry equilibrium, no opportunities for the creation of net surplus for consumers remain unrealised. It also follows, as a theorem in comparative statics, that all freely chosen marginal transactions create weakly positive net increases in surplus for consumers.

Do these results have any normative significance, in the absence of the assumption that preferences are coherent? I suggest they do. They tell us that competitive markets give individuals, moment by moment, what *at the time* they want and are willing to pay for. More precisely: the market mechanism ensures that, at any given moment, if there is some feasible transaction such that everyone would choose to undertake his part of it if it were offered to him, then that transaction takes place. My claim is that this property of the market can sensibly be viewed as worth having, independently of whether

individuals have coherent preferences. Speaking for myself, I would be more definite: even though I have little confidence in the coherence of my own preferences, I believe that this property of the market *is* worth having. My proposal is this: *If CBA is to be interpreted as market simulation, its criterion should be the maximisation of total surplus.*

3. Payment Mechanisms and Context-Dependence

How do we measure surplus? As a starting-point, we have to recognise that all measurements of surplus rely on counterfactual propositions about amounts of money that individuals would or would not pay, or would or would not accept, *in hypothetical situations.* Even in the simplest case of surplus associated with private consumption goods bought in markets, the calculation of surplus depends on propositions about how much consumers would buy if the price of that good was other than it in fact is. But the further we move away from actual markets, the more freedom of manoeuvre we have in specifying the hypothetical *payment mechanism* which we use to define surplus.

For example, suppose we are trying to measure the surplus that individuals gain through having free access to a beach. Since access is currently free, we have to postulate a counterfactual mechanism through which payment could be made. This might be an entry charge levied per visit, or a charge for a permit valid for a specified period. Or, for visitors who come by car, we might consider how the number of visits would change if the price of fuel increased. Or (if we know how to convert time costs into equivalent money costs), we might consider the effects of speed restrictions or of increased congestion on the roads leading to the beach. Or we might translate each individual's choice problem into the domain of collective choice, and imagine that free access is made possible through public expenditure. Then we could consider how each individual would vote in a referendum in which the alternatives were the closure of the beach and the maintenance of free access, financed by an increase in taxation.

If each individual has a single, context-independent preference ordering over outcomes (specified, for example, in terms of visits to the beach, consumption of other goods, time spent travelling, and so on), then the choice of hypothetical payment mechanism is merely a matter of analytical convenience: surplus can be interpreted as a property of preferences, and different payment mechanisms are merely different ways of *eliciting* the same preferences. But if, to the contrary, preferences are not assumed to be coherent, we are not entitled to assume that the surplus created by a transaction is independent of the payment mechanism through which gains from trade are realised. If the surplus attributed to a project depends on which payment

mechanism we postulate, and if any mechanism we postulate is merely hypothetical, what grounds do we have for choosing between mechanisms?

I suggest that this question is the core of the problem of how to deal with anomalies in CBA. The main anomalies which pose problems for CBA can be understood as tendencies for the preferences of given individuals over given outcomes to vary systematically according to the context in which those preferences are elicited or revealed.[7]

For example, the disparity between WTA and WTP is a tendency for an individual's preferences between given combinations of goods and money to vary according to which combination he views as his endowment or reference point. Part-whole bias is a tendency for an individual's valuation of a given element of benefit to vary according to whether that benefit is construed as a free-standing object or as part of some larger entity. Preference reversal is a tendency for an individual's preference ranking of two given lotteries to vary according to whether it is elicited in a dichotomous choice task or in valuation tasks. Asymmetric dominance is a tendency for an individual's ranking of two given options to vary according to whether it is revealed in a choice task in which those are the only two options or in one in which an additional, dominated option is also available. Hyperbolic discounting is a tendency for an individual's ranking of benefits accruing in two given future periods to vary according to the period in which that preference is expressed. In each of these cases, the concept of *the* preference of the individual over the relevant outcomes is ill-defined: preferences are context-dependent. Or, to put this another way, measurements of surplus depend on the form that hypothetical payments are assumed to take.

Some economists like to suggest that conventional theory can easily be adapted to take account of context-independent preferences. To guard against misunderstandings, it is useful to think briefly about how far this claim is valid, and to the extent that it is, what effect it has on my analysis. As an illustration, I consider disparities between WTA and WTP. One possible explanation of these disparities is that individuals' choices and valuations are influenced by reference points. The most fully developed theory of the effect of reference points is probably Tversky and Kahneman's (1991) theory of reference-dependence preferences. Formally, this is a generalisation of the standard Hicksian theory of consumer choice. The generalisation is to define preferences relative to reference points, and to propose particular hypotheses about how rankings of given options change as reference points change. Now one might say that reference-dependence is "only" a new assumption about preferences. Thus, for example, Michael Hanemann (1999, pp. 75–78) acknowledges that reference-dependence may contribute to the WTA/WTP disparities which he has previously explained as the product of income and substitution effects in a conventional Hicksian model. His way of making this concession is to say that reference-dependence "can readily be incorporated

into a standard utility model" by making utility a function, not only of the level of consumption of each good, but also of the reference point for each good. Although Hanemann does not say so, this amounts to a restatement of Tversky and Kahneman's theory. Whether one presents this theory as a radical alternative to the Hicksian theory of consumer choice or as a minor amendment to it is merely rhetoric. What matters is that, contrary to what is assumed in Hicksian consumer theory, preferences over goods vary according to the context in which they are revealed.

Economists sometimes take the further step of re-describing context-dependence as a special kind of preference, with the implication that context is just one of the things that people have preferences over. For example, in a continuation of the passage I quoted above, Hanemann says that, while the reference-dependent utility function "may be more realistic than the standard model ... it can be viewed as a particular case of the standard model with a more specific structure". Thus, while Tversky and Kaheman claim that their theory is a generalisation of the Hicksian theory of consumer choice, Hanemann seems to be claiming the exact opposite. On the face of it, Hanemann is wrong. In reference-dependent theory, there is a family of utility functions $u_r(x)$, where x is the bundle of goods consumed and r is the bundle of goods which serves as the reference point. This is clearly a generalisation of the Hicksian function $u(x)$. Does it make any difference that Hanemann prefers to represent reference-dependence by a single utility function $u(x, r)$? Perhaps what Hanemann means is that "the standard utility model" allows utility to be a function of *any* vector of quantities that matter to the relevant individual, and that $u(x)$ and $u(x, r)$ are equally legitimate as utility functions within this model. But this seems wrong too. On the standard interpretation, utility is a representation of preferences. The hypothesis of reference-dependence is that preferences between given vectors of "goods" *depend on* reference points – not that individuals have preferences *between* reference points. What could it mean to say that an individual prefers, say, "bundle x viewed from reference point r_1" to "bundle x viewed from reference point r_2"? Even if we could somehow imagine a context in which it was meaningful to choose between alternative reference points, we would still need to specify a reference point for *that* choice problem before we could define a preference; we are faced with an infinite regress. The general implication is this: if preferences really are context-dependent, that property cannot be represented in a Hicksian model merely by re-interpreting the objects over which preferences are defined.

So (returning to the main line of argument), the main problem that anomalies pose for CBA can be understood as the context-dependence of preferences. If preferences over goods are context-dependent, measures of surplus for those goods can vary according to the hypothetical payment mechanisms that are used in making the measurements. If this diagnosis is

accepted, the way to cope with anomalies is to find some general, and generally acceptable, criterion for choosing between hypothetical payment mechanisms.

It might seem that the obvious criterion is to look for a payment mechanism that is as close as possible to the one that will actually be used to finance the project under consideration, if it is approved. I shall call this the *actual finance principle*. For example, consider a CBA of a project to build a new highway. Suppose that, if the project is approved, no user charges will be levied, and the costs will be financed from general taxation. Then, if we accept the actual finance principle, we should ask how much each taxpayer is willing to pay in higher taxes for the package of benefits and disbenefits that is generated by the project as a whole.[8] However, this is not the way in which a CBA of a highway project would normally be carried out. A much more common methodology is to estimate the benefits that road-users will gain from the journeys they will actually make, using valuations of travel time savings that have been derived by revealed-preference methods. When this method is used to measure the surplus enjoyed by road-users as a result of a reduction in the time required to make journeys, the hypothetical payment mechanism is not general taxation – even though that is the mechanism that will in fact finance the project. It is whatever form of opportunity cost was used to infer the revealed valuation of time savings. The point of this example is that the actual finance principle, although coherent and apparently natural, is *not* the standard practice in CBA – at least in applications in which revealed-preference valuations are available.[9]

I wish to make a different proposal. This proposal is that, in choosing the hypothetical payment mechanism for any particular good in CBA, we should favour the *closest market analogue*. That is, we should look for the payment mechanism which most closely approximates the mechanisms that are characteristic of competitive markets. By doing this, we ensure that the surplus we are measuring in CBA is as close as possible to the kind of surplus that competitive markets tend to generate. This provides a general strategy for resolving problems caused by anomalies, which is compatible with the interpretation of CBA as market simulation. In the remainder of the paper, I consider some implications of this strategy.

4. Citizens and Consumers

When trying to measure how much an individual is willing to pay for a benefit, or how little she is willing to accept as compensation for incurring a cost, the problem can be framed either as a problem for that individual in isolation, or as a problem for a political community of which she is a member. Take the case of willingness to pay. In the *consumer frame*, we ask

the individual to consider a benefit that accrues only to her, and to say how much she would be willing to pay for that benefit. There is a background assumption that the availability of the benefit to other individuals is independent of her response; and there is a presumption that she is entitled to spend her own income as she chooses. Thus, the individual is being asked: "Given this hypothetical choice problem, to be faced by you as an individual, how would you respond?" In the *citizen frame*, in contrast, we ask the individual to consider a project which confers benefits on many different people; she is just one of those beneficiaries. We postulate some general formula for distributing the costs of the project between individuals – say, an increment on the standard rate of value-added tax. (The formula must be strictly monotonic: the higher the total cost, the more the respondent pays.) The individual is then asked to state the highest cost of the project (either the total cost or, equivalently, the cost imposed on her by the given formula) at which she would prefer it to be undertaken rather than not. Thus, the individual is being asked: "Given this hypothetical problem of collective choice, to be faced by you and your fellow-citizens together, how would you (in the singular) propose that the collectivity responds?"

It is important to notice that the distinction between consumer and citizen frames does *not* correspond with the distinction between private and public goods. According to the standard definition, public goods have two characteristics: non-rivalness in consumption and non-excludability. Non-rivalness is a property of the cost conditions for the supply of a good, and as such has nothing to do with individuals' willingness to pay for that good. Non-excludability is a property of the *actual* technology for extracting payment from consumers, while the discussion of the two frames is conducted in relation to *hypothetical* payment mechanisms.

Thus, valuations of private goods can be elicited in the citizen frame as well as in the consumer frame. Consider surgery for some minor, non-communicable condition (say, varicose veins). In terms of the standard definition of publicness, this is a private good: there is rivalry in consumption, and it is easy to impose charges and to restrict the supply of the good to those who pay. However, in a stated preference survey, we might postulate the payment mechanism that medical care is provided free of charge, that it is rationed according to medical need and length of wait, that the total supply is determined by collective choice, and that the total costs are paid from general taxation. (This is the payment mechanism actually used in the British National Health Service.) Suppose that, for a project which produces a given increase in the total supply of medical care, each respondent is asked to state the highest total cost at which he would prefer the project to go ahead. That is a stated preference exercise which uses the citizen frame. Alternatively, we could use the consumer frame by postulating a hypothetical payment

mechanism by which each individual pays for the specific medical care services he consumes.

Similarly, for many public goods – including some of the "purest" cases of non-rivalness and non-excludability – either frame can be used to elicit valuations. For example, suppose we wish to value the benefits provided by the broadcasts of a public radio station. We could carry out a stated preference survey using a citizen frame by asking each resident of the relevant area to consider a collective project which supplies the broadcasts and which shares the costs between all residents according to some monotonic formula. Each resident would then be asked to state the highest total cost at which he would prefer the project to go ahead. Alternatively, we could use a consumer frame by postulating some hypothetical payment mechanism such that, if the broadcasts are made, each resident can choose independently of the others whether to receive the benefits of the broadcasts in return for the payment of a fee. (Compare the set-top boxes which satellite television companies use as a means of ensuring that their signals can be received only by those who pay to do so.) Each resident would then state the maximum fee he would be willing to pay to receive the broadcasts.

Would the citizen and consumer frames elicit the same valuations? Some explanations of differences between the two types of valuations are consistent with conventional assumptions about preferences. In particular, the goods being valued may have elements of publicness, which are not captured by the payment mechanisms used in the consumer frame. For example, individuals may have altruistic preferences concerning others' consumption of health care.[10] In the case of public broadcasting, individuals may derive benefits from the "conversational value" of having listened to the same programmes as other people have done (Hargreaves Heap, 2002). For the purposes of this paper, however, the most significant cases are those in which differences between the valuations elicited in the two frames are the result of context-dependent preferences. I focus on these cases because, to the extent that they occur, they pose a challenge to any proposed method of dealing with anomalies in CBA. I now consider two forms that such context dependence might take.

First, a respondent may believe that, in the citizen frame, he should express his judgements about what is best for society as a whole, rather than express his private preferences. The idea that collective choice should be understood as the aggregation of private preferences is central to the public choice school of political theory (and, speaking for myself, I think there is much to be said for it). But many people have a different understanding of collective choice: as a process of deliberation about what is best for the collectivity. A respondent who takes this latter position may report a willingness to pay for projects from which he derives no personal benefit, judging that, *from a collective point of view*, the overall benefits outweigh the costs.

For example, responding as a citizen, an able-bodied person might report a willingness to pay his share of the costs of the provision of wheelchair ramps in public buildings. This need not issue from an altruistic preference for the greater welfare of handicapped people. As a private individual, the able-bodied person might not be willing to contribute to a charity which builds wheelchair ramps. He is not saying that *he* wants there to be ramps; he is saying that, in his opinion, *the collectivity* ought to want this.

Second, when a respondent in the citizen frame is presented with a hypothetical cost-sharing formula, she may perceive that formula as unfair. Experimental evidence strongly suggests that many people's willingness to contribute to public goods is governed by ideas of fairness and reciprocity: each person is willing to contribute what she sees as a fair share of the costs of a public good, if and only if other people contribute fairly too. Much of the evidence suggests that negative reciprocity (that is, aversion to paying for other people's free rides, even to the extent of being willing to incur costs to punish free riders) is a stronger motivating force than positive reciprocity (the desire to match other people's contributions).[11] The implication is that, when reasoning as citizens, individuals are sensitive to what they perceive as the fairness or unfairness of the distribution of costs and benefits. Thus, reasoning as a citizen, a respondent might refuse to support a collective project on the grounds that her share of the costs is unfairly high, even though, if reasoning as a consumer, she would have been willing to pay an amount equal to her share of the costs to buy her share of the benefits, and even if, as a private individual, she has no altruistic, malevolent or egalitarian preferences about other people's income or welfare.

If valuations of a given good can be elicited either in the consumer frame or in the citizen frame, and if the two frames can lead to different results, which should be used? According to the general proposal presented in Section 3, we should use whichever is more closely analogous with the payment mechanisms of competitive markets. It is surely obvious that this criterion favours the consumer frame. The market offers choices to individuals, and to groups of individuals (such as economically integrated households) which have already organised themselves to act collectively. It does not provide a framework within which individuals can deliberate about what is best for society as a whole.

So, if CBA is interpreted as market simulation, consumer frames should be preferred to citizen frames as means of eliciting individuals' valuations. In valuing public goods, and in valuing private goods which (like medical care in the National Health Service) are supplied collectively, CBA should try to break down the overall benefits of these goods into benefits that accrue to separate individuals, and then try to find real situations, or to construct hypothetical scenarios, in which individuals choose whether or not to incur private costs to enjoy private benefits. For example, the recreational value of

a beach should be measured by investigating individuals' willingness to incur costs *as individuals* to gain access *as individuals* to the beach, and not their willingness to bear a share of the costs of a public project which allows people in general to have free access to it. Similarly, consider the valuation of a reduction in an environmental health risk, such as exposure to ultraviolet radiation. If the principle of market simulation is accepted, we should not ask respondents to report their willingness to pay taxes to reduce *everyone's* exposure to the risk; instead, we might ask them to report their willingness to pay for a private good (say, sunscreen, or perhaps some hypothetical drug, analogous with hay-fever treatments) which reduces *their own* exposure.

It must be conceded that there are some goods for which credible consumer frames are very difficult to construct. In particular, this is true of public goods that have *existence value*. A thing or event has existence value to an individual if that individual is willing to incur a cost merely to ensure that the thing exists or the event occurs, without his engaging in any act of consumption. Existence value in itself is not incompatible with the consumer frame, as can be seen by considering cases in which existence value is private to one individual. For example, a person might be willing to pay for flowers to be placed on a relative's grave in a distant town, without any intention that these are seen by anyone. If no one else shares the desire that flowers are placed there, the flowers have an existence value that is private to that one person. In such a case, it is clearly possible to have a payment mechanism which uses the consumer frame – that is, a mechanism in which the individual chooses for himself whether or not to buy a good that has value only for him. The problem is to find a consumer frame for *public* existence value. For example, consider the survival of the endangered mountain gorillas of Rwanda. If peace could be restored in Rwanda, there is no doubt that many people would be willing to pay significant amounts of money to *see* these animals in the wild. This is an element of the value of gorilla survival which *can* be valued in a consumer frame. But, I take it, many of the people who support the Dian Fossey Gorilla Fund do so without any intention of visiting Rwanda as eco-tourists: what they value is simply the survival of gorillas. It seems clear that there is existence value here, which people are willing to pay for; but we cannot, even hypothetically, partition this good into private components, supplied separately to different individuals.

If CBA is interpreted as market simulation, we may have to accept that the measurement of public existence value is beyond the scope of CBA. That is not to say that survey methods which use citizen frames to elicit existence values frames are invalid or uninformative, but only that what they measure is not the same thing as the surplus that is measured by market-simulating CBA. I realise that the measurement of public existence value has a flagship status for many environmental economists. Nevertheless, it must be

remembered that this activity is only a small part of the current domain of application of CBA, and even of stated-preference research.

5. When Is Surplus Created?

One major class of anomalies consists of apparent inconsistencies of individuals' preferences across time. Classic examples include failures of self-control and the tendency for people to use hyperbolic rather than exponential discounting. Such patterns of preference violate conventional principles of temporal consistency; they can be understood as cases of context-dependence in which "context" is defined in terms of time. The logic of the market-simulation approach prompts the following question: If individuals' preferences are inconsistent across time, *which* of these conflicting sets of preferences are relevant for competitive markets – and hence for CBA?

In this section, I try to find a general answer to this question. I do so by considering some simple cases in which there could be systematic temporal inconsistencies in individuals' preferences for goods that are traded on markets, and consider how markets would respond to those inconsistencies. Notice that I do not present these cases as direct examples of problems caused for CBA by temporal inconsistency. Rather, I use them to try to tease out the mechanisms by which, when individuals' preferences are temporally inconsistent, competitive markets select which preferences to satisfy. In Section 6, I discuss some concrete implications of this analysis for CBA, given the market-simulation approach.

I begin with an example from everyday life. Suppose that a large class of consumers repeatedly reveal the following dynamic inconsistency.[12] For a consumer who is subject to the effect, it occurs on days on which he plans to spend the evening in a bar. Over the morning and afternoon, he forms a plan to drink moderately. In the evening, after having a few drinks, he gives up the plan, and drinks heavily. Next day, he regrets having drunk so much; this regret helps to motivate him to form another plan for moderate drinking; and so the cycle continues. If we want to measure the benefits that this person derives from his consumption of alcohol, we could refer to the *planning* viewpoint that he takes before he goes into the bar; or to his viewpoint *at the moment of consumption*, in the bar; or to his viewpoint *after the event*, when he reflects on his experience the morning after. But there is no difficulty in saying which of these viewpoints is in fact relevant for the market in alcoholic drinks: the market responds to willingness to pay at the moment of consumption. A bar-owner seeks to gain surplus by offering his customers what they are in fact willing to pay for – not what they previously intended to be willing to pay for, nor what they will later wish they had been willing to pay for. In making forward plans (for example, about what stocks to hold) the

bar-owner acts on *his own* expectations, grounded in his experience of the trade, about what customers will in fact want to buy when the time comes. More generally, competitive markets in consumer goods respond to consumers' willingness to pay at the moment of consumption, as predicted by profit-seeking entrepreneurs; and the market rewards entrepreneurs for the correctness of those predictions.

This conclusion, it must be said, applies only when goods are not produced to meet the specific requirements of individual consumers, and so remain tradable until the point of consumption. For example, compare ready-to-wear clothing with bespoke tailoring. Like the bar-owner, a ready-to-wear clothing manufacturer acts on her own beliefs about what consumers will be willing to pay when her products are put on sale. At the time she commits herself to a production plan, she has no reason to take account of consumers' current beliefs about what they will later be willing to pay (except in so far as this helps her to predict what in fact they will be willing to pay when the goods go on sale). In contrast, a bespoke tailor commits himself to manufacture something that is of specific value to a known individual. Because of the specificity of this commitment, the tailor contracts with the customer before cutting the cloth; in this case, production responds to willingness to pay at the moment the production process becomes specific to the one consumer. I follow the conventions of neoclassical economics in defining an "ideal" or "perfect" market so that, for each good, there are many potential buyers and sellers. Markets in which goods have individual-specific properties are not ideally competitive in this sense.

The principle that competitive markets respond to willingness to pay at the moment of consumption holds true even if there are active futures markets. For example, consider a person's saving behaviour over several years. Suppose her preferences are dynamically inconsistent in the following way. As of now (year 1), her preference is to build up a stock of savings, to be used after year 10, when she will have retired; if she could choose now her mix of consumption and saving for years 1–10, she would choose to save a significant amount each year. However, in year 2, if she has built up savings according to the first year of this plan, she prefers to spend those savings on an expensive holiday. To which of these preferences will markets respond? In this case, unlike those of drinking in bars or buying ready-to-wear clothing, well-developed markets exist for transactions across time. In year 1, the individual can make a contract with a bank, requiring her to make payments over years 1–10 in return for a lump sum to be paid to her in year 11. But if, when year 2 arrives, she wants the holiday, she can undo the effects of her savings plan, simply by borrowing the cost of the holiday from another bank. Thus, the market does not allow her, acting in year 1, to tie her hands in year 2. The amount she saves each year is determined by her preferences *in that year*, within the constraints imposed by her current wealth.

The underlying tendency of markets is to realise gains from trade, defined in terms of surplus *at the last moment at which trade is possible*. In a frictionless market, plans which consumers hold before this last moment can be revised; trades which they make before this last moment can be offset by subsequent trading in the opposite direction. Thus, given sufficient activity by profit-seeking arbitrageurs, prices in periods before the last moment are governed, not by the plans, beliefs and preferences that consumers hold in those periods, but by traders' expectations about the terms on which consumers will be willing to trade when that final moment arrives. Provided that consumer behaviour is predictable in the aggregate, prices are governed by the terms on which consumers will *in fact* be willing to trade at the final moment. In a frictionless competitive market, with no uncertainty,[13] trades can be reversed until consumption begins, but not afterwards.

I suggest that if CBA is to be understood as a simulation of markets, it should follow the same principle. That is, benefits which accrue to individuals in future periods should be assessed in terms of those individuals' valuations *in the periods in which the benefits accrue*, as predicted by the cost–benefit analyst in the light of the best knowledge currently available to her. In making and acting on such predictions, the analyst's role is analogous to that of the entrepreneur in the model of the ideal market. The object of CBA, I maintain, should be to predict the surplus that a project will in fact generate, not to measure what individuals currently believe it to be worth.

6. Coping with Reference-Dependent Preferences

My aim is now to show that the principle proposed in the final paragraph of Section 5 allows CBA to cope with reference-dependent preferences. The argument hinges on the claim that, if surplus is to be measured in terms of individuals' valuations of costs and benefits in the periods in which they accrue, CBA needs to allow for any changes in individuals' perceptions of the costs and benefits of a project that will be induced by the implementation of the project. This is so, whether or not the relevant individuals foresee those changes at the time the CBA is carried out. Thus, if anomalies in ex ante valuations of the costs and benefits of a project are induced by psychological effects which will dissipate quickly if the project is implemented, those anomalies will have correspondingly small effects on a market-simulating CBA.

It is useful to begin by thinking about the distinction between short-run and long-run demand in a case in which issues of anomalies do not arise, and for which CBA has a well-established practice. (Of course, the idea that there are two and only two distinct "runs" is a modelling simplification; my argument does not depend on this particular assumption. What matters is

that, the longer the period of time over which consumers adjust to price changes, the more possibilities there are for substitution.)

For most goods, the price elasticity of demand is less in the short run than in the long. Consider a case in which this effect is due to factors that are recognised by conventional microeconomic theory. As an example, take the demand for water by domestic consumers. If the price of water rises, consumers can respond by substituting less water-intensive technologies for more water-intensive ones – for example, by using washing machines that are more economical in the use of water. Because these technologies are embodied in durable goods, such substitution takes time to complete; hence, the price elasticity of demand for water is less in the short run than in the long. Now consider a CBA of a policy option that increases the price of water. This case is illustrated in Figure 1, which represents the behaviour of a representative consumer. The initial price of water is p_0; if the policy option is adopted, the price will rise to p_1. The short-run demand function for water, given the technology that is optimal for the initial price, is D_0^S. Similarly, D_1^S is the short-run demand function, given the technology that is optimal if the price is p_1. The long-run demand function, D^L, is the locus of price/quantity combinations (p, q) such that, given the technology that is optimal when the price is p, the quantity demanded at that price is q. How should we measure the resulting loss of consumer's surplus?

The standard answer to this question will be familiar to anyone who has studied the principles of CBA. In the short run, the loss of surplus should be defined in terms of D_0^S, the short-run demand function associated with the initial technology; thus, the short-run loss of surplus is measured by the area
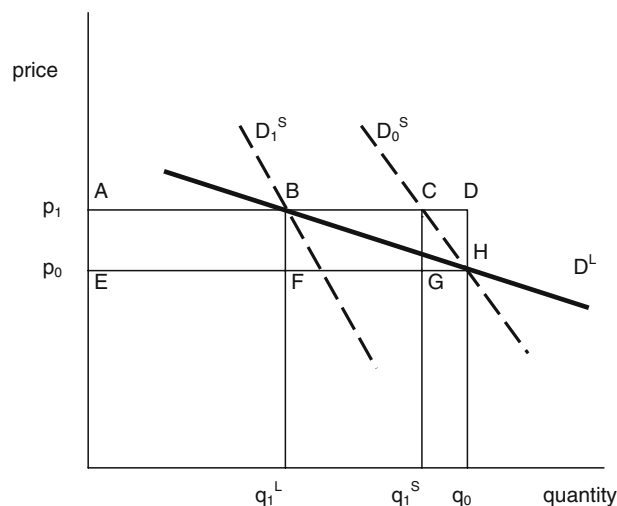

Figure 1. Short-run and long-run demand.

ACHE. But in the long run, the loss of surplus should be defined in terms of long-run demand; thus, the long-run loss of surplus is measured by the area ABHE. Intuitively, this methodology takes account of the fact that, per unit of time, the costs imposed by the price increase are less for a consumer whose capital stock is adapted to the higher price than for one whose capital stock is adapted to the lower price.[14] It would be standard practice in CBA to use this methodology whether or not consumers can accurately predict their own demand functions. In particular, the theoretical argument for using long-run demand functions to measure long-run changes in surplus would not be thought to depend on consumers' ability to predict substitution opportunities. In this respect, CBA standardly takes account of individuals' valuations at the moment of consumption, as predicted by the cost–benefit analyst, and not the valuations that, at the time the analysis is carried out, individuals predict that they will later come to have.

Notice that these measures of changes in surplus are WTA valuations. Viewed in relation to surplus, the distinction between WTP and WTA is no more than the distinction between increases and decreases. If some project induces an increase in an individual's surplus, that increase is her WTP valuation of the benefits of the project. If it induces a decrease in surplus, the absolute value of that decrease is her WTA valuation of the disbenefits of the transaction. In the example, we are measuring a decrease in surplus, and so the measure is of WTA. The general principle is simple: WTP is the measure of gains, and WTA is the measure of losses.[15] Notice, however, that it does not follow from this principle that the stated-preference method which best predicts long-run increases in surplus is *to ask questions about* WTP, or that the method which best predicts long-run decreases in surplus is *to ask questions about* WTA.

Since I am proposing that CBA uses a theoretical framework in which individuals' preferences are not assumed to be coherent, I need to be able to show that this standard treatment of short-run and long-run changes in consumers' surplus is legitimate even if preferences are not coherent, provided only that aggregate demand relationships are predictable and that consumers are price-sensitive. This can indeed be shown. As this is a rather technical issue, it is dealt with in the Appendix.

In the context of this paper, this familiar analysis is significant by virtue of an isomorphism between (on the one hand) the embodied-technology model of short-run and long-run demand and (on the other) the theory of reference-dependent preferences. To explain this isomorphism as simply as possible, I use a restricted form of reference-dependent theory.[16] In fact, the conclusions I reach do not depend on the special assumptions of this particular model; they merely allow me to pass over some complications.

In this model, an individual's preferences *conditional on any given reference point* are described by indifference surfaces with entirely conventional

properties. In particular, these surfaces are smooth and convex everywhere. If reference-dependent preferences have this property, then (given certain technical assumptions), it is possible to define a family of reference-independent indifference surfaces such that, at each point x in goods space, a reference-dependent indifference surface, defined with x as its reference point, is tangent to a reference-independent surface. This configuration is illustrated in Figure 2, in which I_0^S and I_1^S are reference-dependent (or "short-run") indifference surfaces defined in terms of the reference points q_0 and q_1, respectively, and I^L is a reference-independent (or "long-run") indifference surface. Loss aversion is represented by the assumption that short-run indifference surfaces are more convex than long-run ones.

For each individual, a *reference-dependent demand function* can be defined for each reference point q: this shows how the individual's chosen consumption bundle varies with price, given that the reference point remains at q. Provided that reference-independent indifference surfaces are convex, a *reference-independent demand function* can also be defined: this is the locus of price/quantity combinations (p, q) such that, given that the consumer's reference point is q, the quantity demanded at the price p is q. If preferences are loss-averse, reference-dependent demand functions are less elastic than reference-independent ones. This property of the model allows there to be (short-run) disparities between WTA and WTP that are too large to be explained in a standard Hicksian framework. To understand why, recall the well-known result that the ratio of WTA:WTP is greater, the less
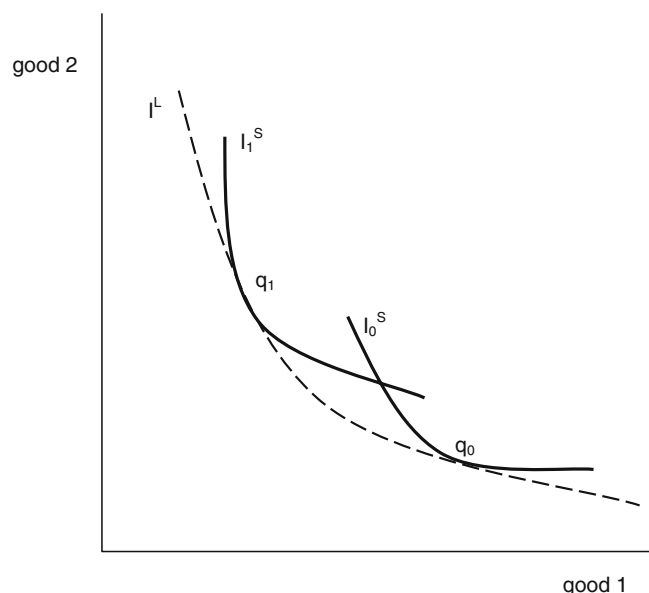

Figure 2. Reference-dependent and reference-independent indifference surfaces.

substitutability there is between money and the relevant good – that is, the more convex indifference surfaces are (Hanemann 1999). However, credible assumptions about the convexity of Hicksian preferences will not induce WTA/WTP disparities of the sizes that are in fact observed (Sugden 1999b). Intuitively, the problem is this: to explain the WTA/WTP disparity as a general phenomenon, we need to postulate that indifference surfaces are very convex in the region of the status quo – wherever the status quo happens to be. In a Hicksian model, in which the status quo has no special status, this requires that indifference surfaces are very convex everywhere, and this assumption has counter-intuitive and empirically false implications. In a reference-dependent model, in contrast, it is coherent to assume that reference-dependent indifference surfaces are extremely convex in the neighbourhood of their respective reference points, while attributing less convexity to underlying preferences, as represented by reference-independent indifference surfaces.

It is easy to see that this model of reference-dependent preferences is isomorphic with the embodied-technology model of short-run and long-run demand. Common to both models is the idea that, as a result of lags in adjustment processes, there is more substitutability between goods in the long run than there is in the short. In one model, the lags result from the embodiment of technology in durable capital. In the other, they result from people's subjective attachment to consumption patterns that have been experienced over the recent past. In each case, the effect of these lags is to make the WTA:WTP ratio greater in the short run than in the long. In the case of reference-dependent preferences, the effect is to induce an "anomalous" value of the short-run WTA:WTP ratio, while the value of the long-run ratio is consistent with Hicksian theory.

As an illustration of this isomorphism, I offer a variant of the example of water consumption. Consider the demand by a representative individual for trips (per period of time) to a particular beach, expressed as a function of the cost of access to that beach. Suppose that the individual's preferences over visits and money take the reference-dependent form I have outlined. Suppose the initial cost of access is p_0 and the number of trips made at that access cost is q_0. Re-interpreting Figure 1, D_0^S is the reference-dependent demand function that is defined in terms of the reference point q_0. D_1^S is the reference-dependent demand function that is defined in terms of the reference point q_1^L. D^L is the reference-independent demand function. Now suppose that, for some extended period of time, the cost of access to the beach has been p_0, and the individual has become accustomed to making q_0 trips. The problem is to evaluate the loss of surplus induced by an increase in the cost of access from p_0 to p_1. If the access cost rises to p_1, the immediate effect (that is, with the reference point remaining unchanged) will be a reduction in the number of trips from q_0 to q_1^S. But if the individual repeatedly consumes a quantity that

differs from her original reference point, there will be a tendency for her reference point to change. This process of adjustment will end only when she reaches a long-run equilibrium in which the quantity she consumes is also her reference point. Thus, the long-run response to the change in access cost is governed by the reference-independent demand function D^L.

Using exactly the same analysis as in the embodied-technology case, the principle that surplus should be measured at the moment of consumption leads to the conclusion that short-run changes in consumers' surplus should be measured by using the reference-dependent demand function that is defined for individuals' initial reference points (in the example, the demand function D_0^S); but long-run changes in surplus should be measured in terms of the reference-independent demand function (in the example, $D^{L)}$. Individuals' initial beliefs about how far (if at all) their reference points will adjust are relevant only in so far as the cost–benefit analyst has reason to expect those beliefs to be reliable.

In considering how to measure changes in surplus when preferences are reference-dependent, we first need to ask: How short is the short run? That is, how quickly do preferences adjust to changes in reference points? The answer to that question determines the relative importance of short run and long-run measures of surplus in CBA.

One of the most striking features of the experimental evidence of reference-dependent preferences is how *easily* people can be induced to change their reference points. A subject who comes into an experimental laboratory and is offered the opportunity to buy, say, chocolates takes her reference point to be the state in which she has no chocolates. But if the same subject is given the chocolates at the start of the experiment, and then a few minutes later is offered the opportunity to sell it, she takes her reference point to include her ownership of the chocolates (Kahneman et al. 1990; Bateman et al. 1997). The implication is that, while people are reluctant to make exchanges which take them away from a perceived status quo, what is perceived as the status quo rapidly adjusts to new circumstances. In addition, evidence is beginning to accumulate which suggests that disparities between WTA and WTP for real transactions in private goods tend to diminish as individuals gain experience of buying and selling the relevant goods (List 2005; Shogren et al. 2001; Loomes et al. 2003). This evidence again suggests that individuals' subjective attachment to reference points is easily disrupted: the special significance of an individual's current holdings as a reference point can be dislodged by manipulations which give salience to the possibility of exchanging those holdings for something else. Because reference points adjust so easily, it has been suggested that loss aversion is a form of myopia – in the sense that individuals' loss aversion, as revealed *ex ante* in decisions, is much greater than can be justified in terms of the *ex post* pain of loss (Kahneman and Sugden 2005; Kahneman and Varey 1991). It seems that the

short run may be very short indeed. In most cases, I suggest, CBA should use measures of *long-run* surplus.[17]

If this conclusion is accepted, we face a second question: What is the best method for measuring long-run surplus? Stated-preference methods which elicit WTP or WTA valuations for postulated policy changes have the demerit that they implicitly ask respondents to evaluate the effects of policy options *as viewed from their current reference points*. Questions which use the WTP format ask respondents to consider hypothetical scenarios in which they incur additional money costs in return for some increase in the supply of a public good. When the WTA format is used, respondents are asked to consider scenarios in which they receive additional money income while experiencing some reduction in the supply of the public good. In each case, the natural assumption is that the respondent is using the current supply of the public good as her reference point. Thus, unless respondents anticipate the tendency for their reference points to adapt to changes in customary consumption,[18] such stated-preference surveys will elicit measures of *short-run* surplus. It is an implication of the theory of reference-dependent preferences that short-run WTP tends to underestimate long-run increases in surplus, while short-run WTA tends to overestimate long-run decreases.

This conclusion provides a reason for favouring revealed-preference methods of preference elicitation to stated-preference methods if both are feasible. Cross-section revealed preference data can used to estimate long-run demand functions directly. For example, the long-run demand to visit recreational sites can be estimated by modelling how the frequency of visits to different sites varies with the cost of access; the long-run demand for savings in travel time can be estimated by modelling how the distribution of trips between travel modes varies with differences in travel time and in money costs.

But what if revealed-preference methods are not feasible? Then we have to look for stated-preference survey instruments which elicit information from which long-run valuations can be inferred. Such survey instruments might be tested and calibrated in relation to goods (for example, travel time) for which reasonably reliable estimates of long-run demand can be derived by revealed-preference methods. Or, as an alternative form of calibration, we might find credible proxies for long-run valuations. For example, List (2005) investigates the effect of experience of buying and selling a good on stated WTA and WTP valuations of that good. If (as List's data suggests) such experience tends to erode WTA/WTP disparities, we might use the responses of the most experienced individuals as estimates of long-run valuations.

One possible research strategy is to compare such imputed long-run valuations of goods with different types of short-run valuations of the same goods, as elicited in stated-preference studies. As a starting point, we might compare WTP, WTA and *equivalent gain* (EG) valuations. (An equivalent

gain valuation of an increase in consumption of a good is the amount of additional money that an individual regards as just as preferable as that increased consumption.) It is a firmly established empirical result that WTA valuations tend to be greater than WTP valuations, and that EG valuations tend to lie between these two extremes. We might ask whether, across a wide range of cases, any one of these valuations is consistently more accurate than the others as a predictor of long-run valuations. The evidence presented by List suggests that long-run WTP is greater than short-run WTP, that long-run WTA is less than short-run WTA, and that the absolute difference between short-run and long-run valuations is much greater for WTA than for WTP. One possible explanation of this latter finding is that, since most people have more experience of buying goods than of selling them, WTP responses reflect more relevant experience than do WTA responses, and so are closer to long-run valuations. An alternative explanation is the hypothesis that anticipated losses of a specific consumption good are more aversive than losses of money, at least in cases in which the loss of money is framed as a payment for a good.[19] Since the EG elicitation task does not confront individuals with either type of loss, there are at least some grounds for the conjecture that EG valuations might be more accurate predictors of long-run valuations than either WTA or WTP (Bateman et al. 1997). But much more theoretical and empirical work will be needed before we can be confident about using short-run stated-preference valuations to predict long-run valuations.

Notes

1. This evidence is reviewed by Camerer (1995) and Sugden (1999a, b). The latter two papers focus on the implications of anomalies for contingent valuation. Readers who want to know my own position on the significance of anomalies may refer to those papers.
2. Conventional economic theory implies that WTA valuations are slightly higher than WTP valuations, as a result of income and substitution effects, but the disparities found in

experimental and survey data are too large to be explained in this way (see Sugden 1999b). Typically, stated valuations increase to some extent with the scale of the relevant good, but the responsiveness of valuation to scale is often too low to be credibly explained by diminishing marginal rates of substitution.

3. For alternative interpretations of the significance of anomalies for economics, see Plott (1996), Binmore (1999), Loewenstein (1999), Starmer (1999), and Bateman et al. (2002, Ch. 12).

4. One of the most important results of social choice theory tells us that, for every (remotely reasonable) method of voting over three or more alternatives, there are profiles of voters' preferences such that some voter has an incentive to misrepresent her preferences (Gibbard 1973).

5. Strictly, a *quantity-compensating variation* or, in the terminology introduced by Hicks (1943), a *compensating surplus*.

6. For example, consider a good which is in fixed supply. Let the total quantity supplied be q and let the initial price be p. Suppose that some individual decides to buy an additional quantity δq (which is small relative to q). This induces a small price increase δp. Buyers of the good lose $q \, \delta p$ in surplus, and there is an equal and opposite gain of surplus by sellers. The value of the transaction which induces this transfer of surplus is $p \, \delta q$. On the neutral assumption that the elasticity of demand is -1, the amount of surplus transferred from buyers to sellers is equal to the value of the transaction itself.

7. The fact that individuals' preferences are subject to stochastic variation creates fewer problems than context-dependence does, because CBA works with aggregated preferences. If stochastic variation is not correlated between individuals, the law of large numbers applies: individual variations cancel out in the aggregate. For stochastic variation to be correlated between individuals, the preferences of different individuals must be conditioned on some commonly observed cue. I interpret such cases as involving context-dependence (the cue being the "context").

8. Notice that there is still a hypothetical element in this payment mechanism, since we need to be able to estimate the highest amount that each person is willing to pay in taxes, and not merely whether or not she is willing to pay the particular tax cost that the project would actually impose on her. This requires that we investigate how people would respond to hypothetical tax rates. If we investigate only people's willingness to pay *actual* tax costs, we do not have the measurements of surplus that are essential for CBA.

9. List (2005) reports that US policy-makers are more willing to trust revealed-preference valuations than stated-preference valuations, other things being equal. The implication is that policy-makers would not endorse the actual finance principle for all forms of CBA.

10. In order for this kind of effect to be capable of being represented in a conventional model, people's other-regarding attitudes must take the form of preferences *over outcomes*. Altruism (preferring that other people consume more rather than less of goods in general, or of specific goods such as health care) and inequality aversion (preferring that the distribution of goods among individuals is more rather than less equal) are attitudes which satisfy this condition. However, some other types of other-regarding attitude, such as a concern for reciprocity, cannot be represented in this way, and so fall into the category of context-dependent preferences. For more on this distinction, see Bardsley and Sugden (2003).

11. The relevant evidence is reviewed in Sugden (1999a). For evidence of individuals' willingness to punish free riders, see Fehr and Gächter (2000).

12. Behavioural theories of choice have been proposed to explain the anomalous patterns of choice associated with failures of self-control and addiction: see Loewenstein and Elster (1992), who discuss a wide range of apparent inconsistencies in individuals' preferences across time.

13. Under uncertainty, trades in state-contingent claims can be reversed up to the moment at which the relevant uncertainty is resolved. The question of how to deal with uncertainty in CBA, given the market simulation approach, will be postponed to another paper.

14. It is tempting but wrong to think that a CBA of the long-run effects of the price increase ought to include as an additional item the costs of the capital goods by which consumers adapt to that increase. All such secondary effects of a price change are already captured in consumers' surplus.

15. In asserting this trusim, I am endorsing a position taken by Knetsch (2005).

16. This special case is examined by Munro and Sugden (2003) in terms of a specific CES (constant elasticity of substitution) functional form for reference-dependent utility. The main conclusions of the analysis which follows can be derived without using the special assumptions of this model – in particular, without assuming that reference-independent indifference curves are smooth everywhere. For an analysis of this issue which does not use these simplifying assumptions, see Sugden (2003).

17. There may be exceptional cases in which reference points adjust slowly. For example, think of the regret that many Californians still feel about the flooding of the Hetch Hetchy valley, approved in 1913. The longevity of regret in this kind of case may reflect the salience of the original reference point, which is perceived as a timeless natural state of the environment. Similarly, perceptions of loss may be long-lived for particular individuals when landscape features which have associations of personal or family memory are destroyed. In cases such as these, the pain associated with movements away from reference points translates into long-lasting losses of surplus, and these losses should be taken into account in CBA.

18. The available experimental evidence suggests that individuals do *not* anticipate the effects of foreseeable changes in their reference points (Loewenstein and Adler 1995).

19. Tversky and Kahneman (1991) offer this hypothesis as an explanation of the observation that, in some experiments they report, EG is closer to WTP than to WTA. The experiments were carried out by Kahneman et al. (1990).

Appendix: The Measurement of Consumers' Surplus When Preferences Are Not Assumed to Be Coherent

In this Appendix, I explain the sense in which the area ACHE in Figure 1 is a measure of the short-run loss of consumer's surplus resulting from an increase in the price from p_0 to p_1, and in which the area ABHE is a measure of the loss of long-run consumer's surplus. The structure of the argument will be familiar to many readers, but it is usually presented in conjunction with assumptions about the coherence of preference. My aim is to show that those assumptions are not necessary.

Consider the short-run response to the price increase. The quantity consumed falls from q_0 to q_1^S. Since consumers' surplus is defined as the excess of willingness to pay over actual payment, the price increase implies a loss of surplus of $q_1^S(p_1-p_0)$ on the transactions that take place at the higher price. On the transactions that would have taken place at the lower price but do not take place at the higher price, the surplus per unit cannot be less than zero and cannot be more than p_1-p_0. Thus, the loss of surplus is bounded below by $q_1^S(p_1-p_0)$, i.e. the area ACGE, and above by $q_0(p_1-p_0)$, i.e. the area ADHE.

By decomposing the price change into a series of smaller increments and by summing changes in surplus over these increments, we can tighten these bounds. Suppose the price rises from p_0 to $p_0+\delta p$, where δp is a small positive increment. The quantity consumed falls to $q_0+\delta q$, where δq is a small negative increment; $(p_0+\delta p, q_0+\delta q)$ is a point on the short-run demand function D_0^S.

The loss of surplus resulting from this incremental price increase is bounded below by $(q_0 + \delta q)\, \delta p$ and above by $q_0\, \delta p$. We can then repeat this process to find lower and upper bounds for the loss of surplus resulting from a further increment in price; and so on. This allows us to calculate upper and lower bounds for the sum of the changes in surplus as the price rises in successive increments from p_0 to p_1. In the limit, as the size of each increment tends to zero, both upper and lower bounds converge on the area ACHE, the area of a strip to the left of the short-run demand curve. A similar argument applied to long-run demand shows that the upper and lower bounds of the loss of long-run surplus as the price rises from p_0 to p_1 converge on the area ABHE, the area of a strip to the left of the long-run demand curve.

The reader may ask whether it is legitimate to add increments of consumers' surplus that result from successive price increments. Two distinct issues arise here. The first concerns income effects. It is a well-known result in Hicksian consumer theory that the compensating variation for a price change is measured by the area of a strip to the left of the *income-compensated* demand curve. In principle, the income-compensated demand function can be defined without reference to utility (and hence, without assuming coherent preferences) by using the Slutsky definition of compensation. In terms of the example: after calculating the loss of surplus for each consumer as a result of the first incremental increase in price (i.e. from p_0 to $p_0 + \delta p$), we subtract this surplus from the consumer's income and find the quantity she would consume at the price $p_0 + \delta p$, given this reduced income. This price/quantity combination is then used as the starting point for calculating the loss of surplus induced by the next incremental price increase; and so on. The price/quantity combinations identified by this procedure trace out an income-compensated demand curve. This method of calculating changes in surplus takes account of the fact that the extraction of surplus from a consumer reduces her income, and hence affects her consumption. However, it is also well-known that, in most CBA applications, the errors created by calculating changes in consumers' surplus from constant-income demand functions rather than from income-compensated functions are tiny. (The fact that these errors are tiny can be verified by investigating the income elasticity of demand. To investigate income elasticity, we do not need to make assumptions about coherent preferences; we merely need to estimate an empirical relationship between quantity demanded and income, with prices held constant).

The second issue concerns path-dependence, and arises only if we are dealing with simultaneous changes in the prices of two or more goods. Conceptually, the change in consumers' surplus caused by a change from one price vector $\mathbf{p}_0$ to another price vector $\mathbf{p}_1$ is calculated by considering a sequence of vectors, each only incrementally different from its predecessor, leading from $\mathbf{p}_0$ to $\mathbf{p}_1$. Are we entitled to assume that the sum of changes of consumers' surplus is independent of the sequence of intermediate price vectors? In Hicksian consumer theory, the official answer is that we are entitled to assume this by virtue of the property of *symmetry of cross-substitution effects*, which is an implication of the standard assumptions about the smoothness of preferences. But, in fact, there is very little evidence that empirically estimated demand functions satisfy symmetry (Deaton and Muellbauer 1980: 60–85). In practical applications of CBA, it is common to sidestep all these problems by using the *rule of half* – that is, to evaluate changes in surplus along a straight-line path from $\mathbf{p}_0$ to $\mathbf{p}_1$, and simply to *assume* that corresponding quantities consumed trace out a straight-line path from the initial quantity vector $\mathbf{q}_0$ to the new vector $\mathbf{q}_1$. This validity (or otherwise) of this rule of thumb seems to be independent of whether preferences are coherent.

References

Bardsley, N. and R. Sugden (forthcoming), 'Human nature and sociality in economics', in S.-C. Kolm and J. Mercier Ythier, eds., *The Economics of Giving, Reciprocity and Altruism*, Amsterdam: North Holland.

Bateman, I., et al. (2002), *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar: Cheltenham.

Bateman, I., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997), 'A Test of the Theory of Reference-Dependent Preferences', *Quarterly Journal of Economics* **112**, 479–505.

Binmore, K. (1999), 'Why Experiment in Economics?', *Economic Journal* **109**, F16–F24.

Camerer, C. (1989), 'An Experimental Test of Several Generalized Utility Theories', *Journal of Risk and Uncertainty* **2**, 61–104.

Camerer, C. (1995), 'Individual Decision Making', in J. H. Kagel and A. E. Roth, eds., *Handbook of Experimental Economics,* Princeton, NJ: Princeton University Press.

Deaton, A. and J. Muellbauer (1980), *Economics and Consumer Behavior*, Cambridge: Cambridge University Press.

Fehr, E. and S. Gächter (2000), 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review* **90**, 980–994.

Gibbard, A. (1973), 'Manipulation of Voting Schemes: A General Result', *Econometrica* **41**, 587–602.

Hanemann, W. M. (1999), 'The Economic Theory of WTP and WTA', in I. Bateman and K. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries,* Oxford: Oxford University Press.

Hargreaves Heap, S. (2002), 'Everybody is Talking About it: Intersubjectivity and the Television Industry', in E. Fullbrook, ed., *Intersubjectivity in Economics,* London: Routledge.

Hey, J. and C. Orme (1994), 'Investigating Generalizations of Expected Utility Theory Using Experimental Data', *Econometrica* **62**, 1291–1326.

Hicks, J. (1943), 'The Four Consumer Surpluses', *Review of Economic Studies* **8**, 108–116.

Kahneman, D. and R. Sugden (2005), 'Experienced Utility as a Standard of Policy Evaluation', *Environmental & Resource Economics* **32**, 161–181 (this issue).

Kahneman, D., J. Knetsch and R. Thaler (1990), 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy* **98**, 1325–1348.

Kahneman, D. and A. Tversky (1979), 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica* **47**, 263–291.

Kahneman, D. and C. Varey (1991), 'Notes on the Psychology of Utility', in J. Elster and J. Roemer, eds., *Interpersonal Comparisons of Well-Being* (pp. 127–163), Cambridge: Cambridge University Press.

Knetsch, J. (2005), 'Gains, Losses, and the USA–EPA Economic Analyses Guidelines: A Hazardous Product?', *Environmental & Resource Economics* **32**, 91–112 (this issue).

List, J. (2005), 'Scientific Numerology, Preference Anomalies, and Environmental Policy-Making', *Environmental & Resource Economics* **32**, 35–53 (this issue).

Loewenstein, G. (1999), 'Experimental Economics from the Viewpoint of Behavioural Economics', *Economic Journal* **109**, F25–F34.

Loewenstein, G. and D. Adler (1995), 'A Bias in the Prediction of Tastes', *Economic Journal* **105**, 929–937.

Loewenstein, G., and J. Elster, eds. (1992), *Choice over Time*, New York: Russell Sage Foundation.

Looomes, G., C. Starmer and R. Sugden (2003), 'Do Anomalies Disappear in Repeated Markets?', *Economic Journal* **113**(2003), C153–C166.

Munro, A. and R. Sugden (2003), 'On the Theory of Reference-Dependent Preferences', *Journal of Economic Behavior and Organization* **50**, 407–428.

Plott, C.R. (1996), 'Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis', in K. J. Arrow, E. Colombatto, M. Perlman

and C. Schmidt, eds., *The Rational Foundations of Economic Behaviour*. Basingstoke: Macmillan.

Shogren, J., S. Cho, C. Koo, J. List, C. Park, P. Polo and R. Wilhelmi (2001), 'Auction Mechanisms and the Measurement of WTP and WTA', *Resource and Energy Economics* **23**, 97–109.

Starmer, C. (1999), 'Experimental Economics: Hard Science or Wasteful Tinkering?', *Economic Journal* **109**, F5–F15.

Starmer, C. and R. Sugden (1989), 'Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses', *Annals of Operations Research* **19**, 79–102.

Sugden, R. (1999a), 'Public Goods and Contingent Valuation', in I. Bateman and K. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries,* Oxford: Oxford University Press.

Sugden, R. (1999b), 'Alternatives to the Neoclassical Theory of Choice', in I. Bateman and K. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries,* Oxford: Oxford University Press.

Sugden, R. (2003), 'Conceptual Foundations of Cost-Benefit Analysis: A Minimalist Account', in A. Pearman, P. Mackie and J. Nellthorp, eds., *Transport Projects, Programmes and Policies: Evaluation Needs and Capabilities,* Ashgate: Aldershot.

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

Tversky, A. and D. Kahneman (1991), 'Loss Aversion in Riskless Choice: A Reference-Dependent Model', *Quarterly Journal of Economics* **106**, 1039–1061.