

Sugden, Robert

**Working Paper**

## Coping with preference anomalies in cost-benefit analysis

CSERGE Working Paper EDM, No. 03-03

**Provided in Cooperation with:**

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

*Suggested Citation:* Sugden, Robert (2003) : Coping with preference anomalies in cost-benefit analysis, CSERGE Working Paper EDM, No. 03-03, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/80272>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**COPING WITH PREFERENCE ANOMALIES  
IN COST-BENEFIT ANALYSIS**

by

**Robert Sugden**

**CSERGE Working Paper EDM 03-03**

**COPING WITH PREFERENCE ANOMALIES  
IN COST-BENEFIT ANALYSIS**

by

**Robert Sugden**

**The Centre for Social and Economic Research  
on the Global Environment  
and**

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

**Contact details:**

**r.sugden@uea.ac.uk; tel: (44)(0) 1603 593423**

**Acknowledgements:**

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the interdisciplinary research programme of the ESRC Centre for Social and Economic Research on the Global Environment (CSERGE).

This paper was written as part of the Programme in Environmental Decision Making, organised through the Centre for Social and Economic Research on the Global Environment, and supported by the Economic and Social Research Council of the UK (award no. M 535 25 5117). My work was also supported by the Leverhulme Trust. An earlier version of the paper was presented at the Second World Congress of Environmental and Resource Economists at Monterey, California in June 2002. In writing the paper, I have benefited from discussions with and comments from Ian Bateman, Robin Cubitt, Jack Knetsch, John List, Graham Loomes, Charles Mason, Alistair Munro and Shepley Orr.

**ISSN 0967-8875**

## Abstract

Economics conventionally assumes that preferences are *coherent*, *i.e.* stable, context-independent, and consistent with axioms of rationality. Since these assumptions underpin standard interpretations of cost-benefit analysis (CBA), preference ‘anomalies’ found in stated preference surveys pose severe problems. This paper proposes an alternative interpretation of CBA as the simulation of the workings of competitive markets in situations in which, in reality, there is market failure. The object of a CBA of a project is to measure the surplus that is created by that project. It is shown that surplus can be defined without assuming that individuals have coherent preferences. Some of the ways in which this approach to CBA can overcome problems caused by preference anomalies are discussed.

**Key words:** Cost-benefit analysis, anomalies, surplus, willingness to pay.

## 1. Introduction

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, invoke 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. These anomalies pose severe problems for normative economics. In CBA, the problems are particularly acute when anomalies show up in the survey data from which preferences are to be inferred. Unfortunately, this is often the case. In environmental economics, particularly intractable problems have been caused by disparities between willingness-to-accept (WTA) and willingness-to-pay (WTP) valuations and by the insensitivity of stated valuations to the scale of the good being valued.<sup>1</sup>

In this paper, I propose an approach to CBA which uncouples it from the assumption of preference coherence. If all individuals’ preferences are coherent, this proposal affects only the interpretation of CBA: given that assumption, any measurements of net benefit that are valid in terms of my proposal are also valid in terms of current CBA methodology. However, my proposal allows meaningful measurements of cost and benefit to be made even if preferences are not coherent.

---

<sup>1</sup> This evidence is reviewed by Camerer (1995) and Sugden (1999a, 1999b). For alternative interpretations of the significance of anomalies for economics, see Plott (1996), Binmore (1999), Loewenstein (1999), and Starmer (1999).

## 2. 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. For example, preference data may be interpreted as evidence about how, under ideal conditions, individuals would vote in single-issue referenda. The *market simulation* strategy sees the purpose of CBA as to simulate the workings of an ideal competitive market, given the presence of public goods, externalities and other forms of market failure.

If the conventional assumptions about preferences are valid, these three rhetorical strategies are mutually consistent. If each individual has just one preference ordering, which governs her behaviour both in markets and in the political process, then the welfarist and public choice approaches coincide. From the welfarist perspective, these preferences can be interpreted as measures of the individual's welfare, while from the public choice perspective they can be interpreted as the judgements she makes as a citizen. It is an implication of the first fundamental theorem of welfare economics that simulating the equilibrium state of an ideal competitive market implies achieving Pareto efficiency, which is a necessary condition for the maximisation of social welfare as conventionally defined.

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 shall propose a form of CBA which, even if preferences are not coherent, can 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 *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 prescribe which option should be chosen. What is necessary is that the information 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 3, 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. Independently of the relative merits of different CBA methodologies, there is some value in standardisation. 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 conflicting 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 supranational); 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 government seeks co-funding for a project from another tier, or when a private firm seeks co-funding from government, a 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 that 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.

### 3. 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*.<sup>2</sup> 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, negative or zero) 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

---

<sup>2</sup> Strictly, a *quantity-compensating variation* or, in the terminology introduced by Hicks (1943), a *compensating surplus*.

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 preferences, 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.

Conventional welfare economics analyses the capacity of 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 a 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. 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  $\delta q$  (which is small relative to  $q$ ). This induces a small price increase  $\delta p$ . 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.

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 together. 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, surplus can be defined without using that assumption. One might then ask whether the results of conventional welfare economics still apply: do competitive markets realise opportunities for the creation of net surplus, even if individuals' preferences are not coherent? In another paper (Sugden, 2002b), 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 actor: *consumers* and *traders*. Consumers come to the market with endowments of various 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' preferences is 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, such that 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.* Recall that I am not proposing that this criterion ought to govern public decisions, but only that it should be the organising principle for the accounting conventions of CBA.

#### 4. Payment Mechanisms

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. Suppose we treat these individuals as if they were on the buying side of a notional market. Then, for each individual, we want to measure the maximum she would be willing to pay for beach access. 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 there were two options: the closure of the beach, and the maintenance of free access, financed by an increase in taxation.

*If* individuals' preferences satisfy the standard conditions of coherence, 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 we abandon the assumption that preferences are coherent and instead think in terms of surplus, 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 measures of surplus differ according to the payment mechanism used, that does not invalidate the concept of surplus: it is simply a fact about the economic world. In understanding how markets work, we need to take account of the fact that the design of payment mechanisms is one of the many ways in which entrepreneurs seek to appropriate surplus. Just as sellers of goods try to buy or

produce those goods at the lowest cost, and just as they try to make these goods as attractive as possible to potential buyers, so they also try to find payment mechanisms that will bring in as much revenue as possible. And just as competition limits the power of any seller to appropriate surplus through charging high prices, so it limits sellers' power to appropriate surplus through *any* payment mechanism. The driving force of a market economy comes from the attempts of competing traders to seek and to appropriate surplus. The same wealth-seeking, competitive dynamic that leads firms to minimise production costs and to give their products the characteristics that consumers are willing to pay for also induces them to seek out the most profitable payment mechanisms for selling those products. In a perspective in which surplus, rather than preference, is the fundamental concept, it is artificial to try to distinguish between these different ways of creating surplus. Surplus must be defined in relation to specific payment mechanisms.

The problem for CBA is that that its measurements of surplus depend on *hypothetical* payment mechanisms. 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? The difficulties that preference anomalies cause for CBA can be interpreted as manifestations of this general problem. 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.

My 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 my proposed interpretation of CBA. In the rest of the paper, I suggest some concrete ways in which this strategy might be applied.

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

Similarly, for most public goods, either frame can be used to elicit valuations. (A special type of public good for which this may not be true will be discussed later). For example, consider a paradigm case: a project to build a lighthouse. Let us assume that the benefits of the lighthouse will accrue entirely to the members of some given population of recreational sailors. We could carry out a stated preference survey using a citizen frame by asking each sailor to consider a collective project which builds the lighthouse and which shares the costs between all sailors according to some strictly monotonic formula. Each sailor 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 lighthouse was built, each sailor could choose independently of the others whether to receive the benefits of the lighthouse in return for 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 sailor would then state the maximum fee he would be willing to pay to receive the benefits of the lighthouse, if it was built.

If conventional assumptions about preferences are satisfied, the consumer and citizen frames will elicit the same valuation of any given benefit by a given individual. More specifically, the same valuation will be elicited *if* the individual has coherent preferences over combinations of money and benefit accruing to her, and *if* she treats those private preferences as the appropriate basis for making proposals about how the collectivity should choose. From the viewpoint of a theory which assumes that these two conditions are satisfied, systematic differences between the valuations elicited in the two frames are anomalies.

However, anomalies may occur in stated preference studies as a result of individuals' *not* using their private preferences as the basis for their responses in the citizen frame. Here are three forms that such anomalies 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. Notice that this need not be the same thing as 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).<sup>3</sup> 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 – and are perhaps predisposed to interpret the requirements of fairness in self-serving ways. 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.

Third, if people interpret contributions to public goods in terms of the fair sharing of costs and benefits, contributions to different goods may be treated as close substitutes, even if the goods themselves are not. For example, a person who values the work done by conservation charities in general may perceive an obligation of fairness to make some contribution to them. But, he may reasonably think, this obligation can be discharged by making a sufficient total contribution to some subset of conservation charities; if different people can be relied on to support different charities, it is not necessary that *he* contributes to every charity whose work he values. Thus, the amount that a person is willing to pay as a

<sup>3</sup> The relevant evidence is reviewed in Sugden (1999a). For evidence of individuals' willingness to punish free riders, see Fehr and Gächter (2000).

voluntary contribution towards the supply of a particular public good need not be any kind of measure of his valuation of the benefits he derives from it. This form of substitutability between contributions may be a partial explanation of the scale insensitivity (or ‘embedding effect’, or ‘part-whole bias’) found in stated preference studies in which payments are described as voluntary contributions (e.g. Kahneman and Knetsch, 1992).<sup>4</sup>

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 4, 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.

I have argued that the driving force of markets is the motive to realise and appropriate surplus. Because it is difficult to appropriate surplus from collective consumption, profit-seeking traders try to find ways of privatising what would otherwise be public goods. Thus, we find payment mechanisms, such as pay-as-you-view television, enclosed sports grounds, health clubs, private beaches and so on, which are designed to separate out those private consumption components of jointly-consumed goods that individuals can be induced to pay for, and to restrict enjoyment of benefits to people who pay. We might say that the dynamic of the market is constantly directed towards the discovery and propagation of consumer frames.<sup>5</sup>

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

<sup>4</sup> Scale insensitivity cannot be attributed *entirely* to respondents’ attitudes to fairness, since a qualitatively similar effect is found, in a much weaker form, in experimental subjects’ valuations of private consumption goods (Bateman *et al.*, 1997b).

<sup>5</sup> Entrepreneurs occasionally find ways of using citizen frames to tap surplus. For example, the use of ‘environmentally friendly’ characteristics of products as selling points might be interpreted as a way of appropriating surplus from individuals’ willingness to participate in a collective enterprise of conservations.

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.<sup>6</sup>

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 (a charitable trust which works to protect mountain gorillas) 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. It seems that, in cases of pure public existence value, there may be no alternative to the use of citizen frames: in the absence of anything better, these may be the closest analogues to the market we can find.

<sup>6</sup> Notice that the person who pays for the flowers may be acting on a sense of moral duty. As this example shows, valuations expressed in the consumer frame are not necessarily self-interested.

## 6. When is Surplus Created?

One major class of anomalies which impinge on CBA consists of apparent inconsistencies of individuals' preferences across time. In addition to anomalies that are direct violations of conventional principles of dynamic consistency, such as failures of self-control and the tendency for people to use hyperbolic rather than exponential discounting,<sup>7</sup> there is (as I shall explain later) some reason to think that the well-known tendency for people to prefer to maintain what they perceive as the status quo rather than to move to alternative states is a form of myopia. The logic of the market-simulation approach to CBA prompts the following question: If individuals' preferences are inconsistent across time, *which* of these conflicting sets of preferences are relevant for markets?

Consider an example from everyday life. Suppose that a large class of consumers repeatedly reveal the following dynamic inconsistency.<sup>8</sup> For a consumer who is subject to the effect, it occurs on days in 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 next day. 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.

The principle that the market responds to willingness to pay at the moment of consumption holds true even if there are active futures markets. Consider a two-period economy in which some good can be consumed only in the second period, but claims on this good can be traded against money in both periods. Suppose there is a predictable tendency for consumers in period 1 to under-

<sup>7</sup> Loewenstein and Elster (1992) review the anomalies that have been found in individuals' choices over time.

<sup>8</sup> 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).

estimate the extent to which they will want the good in period 2. Then, in the absence of arbitrage, the price of the good will be higher in period 2 than in period 1: the price in period 2 reflects what consumers are willing to pay at the moment of consumption, while the price in period 1 reflects what they are willing to pay in advance. But if the period 2 price can be predicted by traders, any difference between the two prices presents an arbitrage opportunity, which will be eliminated by the surplus-seeking activities of traders. In an equilibrium in which all arbitrage opportunities have been exploited, the price in *both* periods reflects what consumers are in fact willing to pay at the moment of consumption.

As a more concrete example of the workings of markets through time, 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 is retired; if she could choose now her mix of consumption and saving for years 1 to 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 that of drinking in bars, well-developed markets exist for transactions across time. In year 1, she can make a contract with a bank, requiring her to make payments over years 1 to 10 in return for a lump sum to be paid 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. In a frictionless competitive market, with no uncertainty,<sup>9</sup> trades can be reversed until consumption begins, but not afterwards.

<sup>9</sup> Under uncertainty, trades in state-contingent claims can be reversed up to the moment at which the relevant uncertainty is resolved.

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 valued in terms of 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.<sup>10</sup> The object of CBA should be to predict the surplus that a project will in fact generate, not to measure what individuals currently believe it to be worth. I now consider some implications of this principle.

---

<sup>10</sup> In terms of the analogy with ideal markets, the cost-benefit analyst assumes that entrepreneurs in such a market act on the same 'best available knowledge' as she herself has access to.

## 7. The Significance of Long-Run Demand

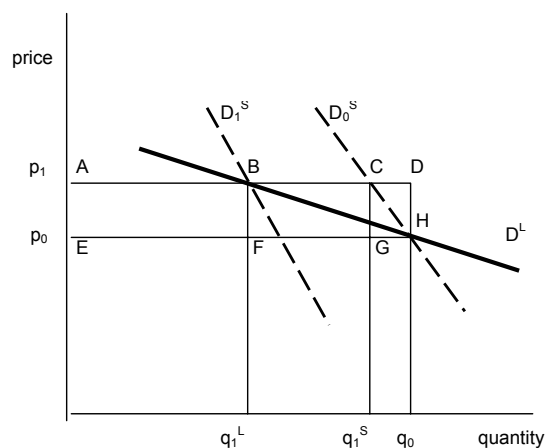
My aim is now to show that the principle proposed in Section 5 allows CBA to cope with some significant anomalies. The argument hinges on the distinction between short-run and long-run demand. It is useful to begin by thinking about this distinction in a case in which issues of anomalies do not arise, and for which CBA has a well-established practice.

For most goods, the price elasticity of demand is less in the short run than in the long. Consider a case in which this phenomenon is due to factors that are recognised by conventional microeconomic theory: 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. But 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  $q$ . In calculating the resulting loss of consumer's surplus, which demand function should we use?

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 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 subjective costs imposed by the price increase are greater for a consumer whose capital stock is adapted to the lower price than for one whose capital stock is adapted to the higher price. 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 individuals predict themselves to have.

**Figure 1: Short-run and long-run demand**



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 in the absence of such assumptions, provided only that consumers are price-sensitive. This can indeed be shown. As this is a rather technical issue, it is dealt with in Appendix 1.

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*, which I now outline. According to this theory, an individual's preferences over given options vary according to which bundle of consumption goods she treats as her default position or *reference point*. Given the hypothesis that losses (defined relative to reference points) are more aversive than equal and opposite gains are attractive, this theory can explain disparities between WTA and WTP (Tversky and Kahneman, 1991; Munro and Sugden, 2003). Reference-dependence may also be implicated in various anomalies in individuals' relative valuations of different quantities of gains or of losses. According to the

hypothesis of *diminishing sensitivity*, small gains are disproportionately attractive relative to large gains, and small losses are disproportionately aversive relative to large losses. This effect may contribute to the scale insensitivity anomaly discussed in Section 5. Transport economists have found a scale anomaly which works in the opposite direction: the tendency for survey respondents to report very low per-minute WTP valuations for savings in travel time when the savings being offered are themselves small. In this case, it seems that respondents are treating very small quantities of a benefit as too insignificant to bother about, in much the same way that experimental subjects sometimes treat very small positive probabilities as if they were zero.<sup>11</sup>

The premise of the present paper is that CBA should be uncoupled from the standard assumption that individuals have coherent preferences. It would be contrary to the spirit of that premise to substitute a different set of restrictive (if non-standard) assumptions about preferences. Ultimately, then, I want to propose a way of dealing with reference-dependence which does not depend on highly-specific assumptions. As a starting point, however, I shall use a special form of reference-dependent preference theory, in which the isomorphism with the embodied-technology model is particularly obvious. The implications of relaxing these special assumptions are considered later.

In the special form of the theory, 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.<sup>12</sup>

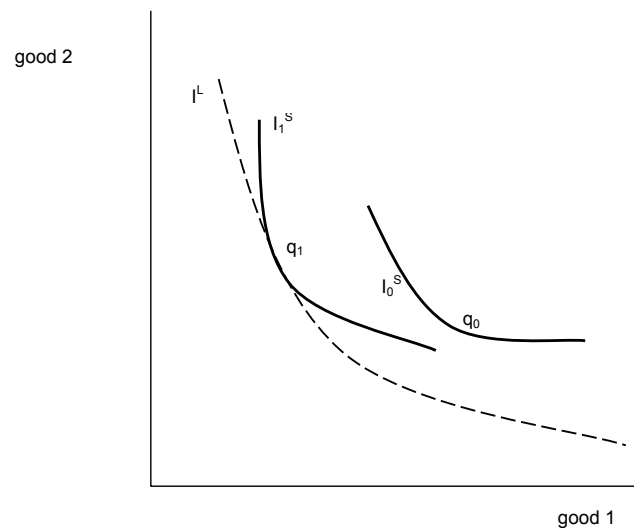
For each individual, a *reference-dependent demand function* can be defined for each reference point  $q$ : this shows how the individual's chosen consumption bundle varies with price, given that the reference point remains at  $q$ . Provided that reference-independent indifference surfaces are convex, a *reference-independent demand function* can also be defined: this is the locus of

<sup>11</sup> That small positive probabilities are sometimes rounded to zero is one of the editing operations proposed by Kahneman and Tversky (1979).

<sup>12</sup> 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.

price/quantity combinations  $(p, q)$  such that, given that the consumer's reference point is  $q$ , the quantity demanded at the price  $p$  is  $q$ . In a model of this kind, the usual disparity between WTA and WTP is an implication of the assumption that reference-dependent indifference surfaces are more convex than reference-independent ones (as in the case illustrated in Figure 2). This disparity is one manifestation of a more general property of the model, that reference-dependent demand functions are less elastic than reference-independent ones.

**Figure 2: Reference-dependent and reference-independent indifference surfaces**



It is now easy to see that this model of reference-dependent preferences is isomorphic with the previous 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 the first model, the lags result from the embodiment of technology in durable capital. In the reference-dependent model, they result from people's subjective attachment to consumption patterns that have been experienced over the recent past.

As an illustration of this isomorphism, I offer a variant of the example of water consumption. Consider the demand by a representative individual for visits (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 visits 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  $q_0$  visits. 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 following conclusions. 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.

These conclusions have significant implications for the way that CBA deals with anomalies that can be traced to reference-dependence. In so far as they impinge on CBA, these anomalies usually show up in data generated by stated-preference survey methods. Usually, such surveys 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,<sup>13</sup> such stated-preference surveys will elicit measures of *short-run* surplus. Whether short-run or long-run surplus is more relevant for CBA depends on how quickly reference points adjust, relative to the life of a project.

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, some 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 them, she takes her reference point to include her ownership of the chocolates (Kahneman, Knetsch and Thaler, 1990; Bateman *et al.*, 1997a). 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 (Shogren *et al.*, 2001; List [this volume]; Loomes, Starmer and Sugden, 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 Varey, 1991; Kahneman [this issue]).

Taken together, this body of evidence suggests that the short run may be very short indeed. If that is the case, stated-preference surveys elicit measures of the wrong kind of surplus for use in CBA. What we need are measures of *long-run* surplus.

<sup>13</sup> The available experimental evidence suggests that individuals do *not* anticipate the effects of foreseeable changes in their reference points (Loewenstein and Adler, 1995).

More fundamentally, we need to know much more than we currently do about the rates at which reference points adjust for different kinds of goods. There may be exceptional cases in which reference points adjust slowly; in such cases, CBA needs measures of short-run as well as long-run surplus.<sup>14</sup> However, my conjecture is that, for most policy options that are subjected to CBA, reference points adjust sufficiently quickly that the surplus created by an option over its life can be adequately approximated by using long-run valuations.

Long-run surplus can be measured most directly by using cross-section revealed-preference data. 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. Provided that the relevant cross-section studies generate smooth demand functions, calculations of long-run changes in surplus will not be subject to discontinuities between WTA and WTP: relative to any given level of initial consumption, the *long-run* valuation of marginal increases in consumption will be equal to that of marginal decreases.

Up to now, I have been using restrictive assumptions about reference-dependent preferences which imply the existence of smooth long-run demand functions. What if those assumptions are relaxed? As I show in Appendix 1, the legitimacy of using demand functions to measure consumers' surplus does not depend on any specific theoretical explanation of those functions (other than the general assumption that individuals are price-sensitive). So the crucial question is whether or not consumers' aggregated responses to prices can be represented by smooth long-run demand functions. Ultimately, that is an empirical issue.<sup>15</sup>

<sup>14</sup> For example, think of the regret that many Californians still feel about the flooding of the Hetch-Hetchy valley, approved in 1913, or that Australians still feel about the extinction of the Tasmanian wolf, which is presumed to have occurred in the early twentieth century. The longevity of regret in these cases may reflect the salience of the original reference points, which are perceived as timeless natural states 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.

<sup>15</sup> Practitioners of CBA generally recognise that, in order for consumers' surplus measurements to be valid, it is sufficient to be able to make accurate predictions of demand responses to price changes; it is not necessary that the demand models that generate these predictions can be derived from specific assumptions about rational choice. For example, when CBA is used in transport planning, it is standard practice to calculate consumers' surplus using traffic-generation, modal-split and route-assignment models which are not derived from rationality assumptions.

The question we need to ask is whether, given a body of relevant cross-section data concerning prices and quantities (for example, cross-section data about access costs and numbers of visits for different recreational sites), a smooth demand function can be fitted econometrically. Experience seems to suggest that such functions *can* normally be fitted to relevant cross-section data. In Appendix 2 I discuss possible explanations of this (supposed) fact, which are compatible with short-run disparities between WTA and WTP, and which do not depend on restrictive assumptions about the form taken by reference-dependent preferences.

Unfortunately, it is not always possible to find revealed-preference data from which long-run demand relationships can be estimated. Thus, I suggest, an important task for stated-preference research is to find 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 there already exist reasonably reliable estimates of long-run demand, derived by revealed-preference methods.

For example, one possible research strategy is to compare long-run valuations of goods, as inferred from revealed-preference data, 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 short-run valuations is consistently more accurate than the others as a predictor of long-run valuations, or whether long-run valuations can be predicted by using some general formula which combines two or all of WTP, WTA and EG.

The evidence presented by List ([this issue]) suggests that long-run valuations are understated by short-run WTP and overstated by 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 finding is that loss aversion is a short-lived psychological effect, and 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).<sup>16</sup> Since the EG elicitation task does not confront individuals with either type of loss,

<sup>16</sup> Kahneman, Knetsch and Thaler (1990) offer the latter hypothesis as an explanation of the observation that, in the experiments they report, EG is closer to WTP than to WTA. (They do not discuss whether loss aversion is short-lived).

there are at least some grounds for the conjecture that EG valuations are more accurate predictors of long-run valuations than either WTA or WTP.<sup>17</sup> 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.

## 6. Conclusion

In this paper, I have defended an interpretation of CBA as the simulation of the workings of competitive markets in situations in which, in reality, there is market failure. On this interpretation, the object of a CBA of a project is to measure the surplus that is created by that project. I have argued that surplus can be defined without assuming, as conventional welfare economics does, that individuals have stable, context-independent and internally consistent preferences. Thus, this methodology gives us a way of appraising projects in a way that does not require those questionable assumptions. I have suggested that, by following this methodology, we may be able to avoid some of the problems that preference anomalies have been thought to cause for cost-benefit analysis.

<sup>17</sup> This conjecture is presented by Bateman *et al.*, (1997a).



## References

- Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997a). A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics* 112: 479-505.
- Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997b). Does part-whole bias exist? An experimental investigation. *Economic Journal* 107: 322-332.
- Binmore, K. (1999). Why experiment in economics? *Economic Journal* 109: F16-24.
- Camerer, C. (1995). Individual decision making. In John H. Kagel and Alvin E. Roth (eds), *Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.
- Deaton, A., and Muellbauer, J., (1980). *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.
- Fehr, E., and Gächter, S., (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90: 980-994.
- Hicks, J., (1943). The four consumer surpluses. *Review of Economic Studies* 8: 108-116.
- Kahneman, D., [Paper to be submitted for special issue of *Environmental and Resource Economics*].
- Kahneman, D., and Knetsch, J., (1992). The purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22: 57-70.
- Kahneman, D., Knetsch, J., and Thaler, R., (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy* 98: 1325-1348.
- Kahneman, D., and Tversky, A., (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263-291.
- Kahneman, D., and Varey, C., (1991). Notes on the psychology of utility. In Jon Elster and John Roemer (eds), *Interpersonal Comparisons of Well-Being*. Cambridge: Cambridge University Press, pp. 127-163.
- List, J., [Paper to be submitted for special issue of *Environmental and Resource Economics*].
- Loewenstein, G., (1999). Experimental economics from the viewpoint of behavioural economics. *Economic Journal* 109: F25-34.
- Loewenstein, G., and Adler, D., (1995). A bias in the prediction of tastes. *Economic Journal* 105: 929-937.
- Loewenstein, G., and Elster, J., (eds) (1992). *Choice over Time*. New York: Russell Sage Foundation.
- Loomes, G., Starmer, C., and Sugden, R., (2003). Do anomalies disappear in repeated markets? Forthcoming in *Economic Journal*.

- Munro, A., and Sugden, R., (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 Kenneth J. Arrow, Enrico Colomatto, Mark Perlman and Christian Schmidt (eds), *The Rational Foundations of Economic Behaviour*. Basingstoke: Macmillan.
- Shogren, J., Cho, S., Koo, C., List, J., Park, C., Polo, P. and Wilhelmi, R. (2001). 'Auction mechanisms and the measurement of WTP and WTA.' *Resource and Energy Economics*, vol. 23, pp. 97-109.
- Starmer, C., (1999). Experimental economics: hard science or wasteful tinkering? *Economic Journal* 109: F5-15.
- Sugden, R., (1999a). Public goods and contingent valuation. In Ian Bateman and Kenneth 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 Ian Bateman and Kenneth 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., (2002a). Conceptual foundations of cost-benefit analysis: a minimalist account. Forthcoming in Alan Pearman, Peter Mackie, John Nellthorp and Liana Giorgi (eds), *Transport Projects, Programmes and Policies: Evaluation Needs and Capabilities*, Ashgate.
- Sugden, R., (2002b). The responsibility criterion: consumer sovereignty without the assumption of coherent preferences. Mimeo.
- Tversky, A., and Kahneman, D., (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics* 106: 1039-1061.

## Appendix 1: 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 consumers' surplus resulting from an increase in the price from  $p_0$  to  $p_1$ . (A similar argument shows the sense in which the area ABHE is a measure of the loss of long-run consumers' surplus.)

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  $\delta p$  is a small positive increment. The quantity consumed falls to  $q_0 + \delta q$ , where  $\delta q$  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.

## Appendix 2: Long-run Demand Functions

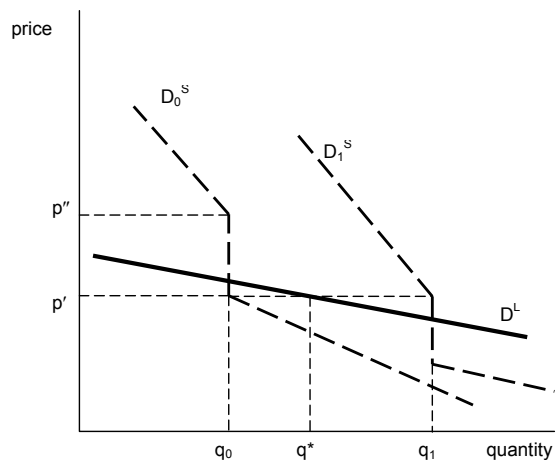
Formally, a long-run demand function is a locus of price/quantity pairs  $(p, q)$  for some population of consumers, such that each pair is *stable* in the following sense: if the price remains at  $p$  for a sufficiently long time, the quantity consumed will converge to  $q$ . Implicit in this definition (by virtue of the mathematical meaning of the concept of a 'function'), and essential for the measurement of surplus, is the assumption that, for any given price  $p$ , there is *one and only one* quantity  $q$  such that  $(p, q)$  has this stability property. In other words, it is assumed that there is some mechanism which, at any given price, induces convergence to some unique quantity.

At first sight, it may seem that the existence of such a mechanism is incompatible with the version of reference-dependent preference theory proposed by Tversky and Kahneman (1991). In this version of the theory, reference-dependent preferences defined in relation to any given reference point in goods space have an elasticity of substitution of zero at that point. (Thus, reference-dependent indifference surfaces are kinked at the relevant reference points.) The implications of this assumption for reference-dependent demand functions are shown in Figure 3. For any individual, for any given reference point (interpreted as a level of customary consumption of the relevant good), there is a range of prices over which that individual's reference-dependent demand function is totally inelastic. Thus, an individual's reference-dependent demand functions, such as  $D_0^S$  and  $D_1^S$ , are 'stepped' at their respective reference points. This property of demand functions is consistent with the common observation that the ratio between the WTP valuation of a given increment of consumption and the WTA valuation of an equal and opposite decrement remains significantly less than 1, even when the size of that increment is very small. (In the case shown in Figure 3: if the reference point is  $q_0$ , the WTP valuation of small increments of consumption above  $q_0$  is  $p'$  per unit, while the WTA valuation of small decrements below  $q_0$  is  $p''$ .) If reference-dependent demand functions take this form, there is for each price  $p$  a range of values of  $q$  such that each  $(p, q)$  lies on the reference-dependent demand function defined for the reference point  $q$ . (In Figure 3: if the price is  $p'$ , this range extends from  $q_0$  to  $q_1$ .) Thus, Tversky and Kahneman's theory does not identify a unique reference-independent quantity for each price.

However, this is not to say that *there cannot be* a process of convergence to a long-run demand function – only that Tversky and Kahneman's model *does not include* any such mechanism. Under certain additional assumptions, convergence to a long-run demand function *does* occur. I now present a simple

model of such a mechanism. This is intended merely as an illustration of the modelling possibilities.<sup>18</sup>

**Figure 3: Discontinuous short-run demand**



For the purposes of the model, I assume that there is a sequence of discrete time periods. There is a population of consumers, each of whom is active in the market in each period. Each consumer has reference-dependent preferences over combinations of current consumption of the relevant good and ‘money’ (standing for current consumption of all other goods). These preferences are assumed to be the same for all consumers and all time periods. However, different consumers may have different reference points, and a consumer’s reference point can change over time. Specifically, in each period  $t$ , each consumer’s reference point is given by her consumption of the relevant good in period  $t - 1$ . Given this reference point, a reference-dependent demand function can be defined for each consumer in each period. However, there is some noise in the model: if, in any given period, a consumer’s reference-dependent demand function implies that the quantity consumed at the prevailing price should be  $q'$ , actual consumption is  $q' + \varepsilon(q')$ , where  $\varepsilon(q')$  is a random variable associated

<sup>18</sup> An alternative model, based on the assumption that there is some turnover of consumers over time, is presented in Sugden (2002a).

with the quantity  $q'$ , independently distributed for each individual.<sup>19</sup> If the random process which generates this disturbance term is specified, and if the price is held constant, these assumptions define a dynamic random process by which demand evolves over time from any given initial profile of consumption quantities (i.e. one such quantity for each consumer in the population). Under many alternative assumptions, this process will converge to a unique level of expected consumption per consumer, independent of the initial consumption profile. Hence, given a sufficiently large number of consumers, there is convergence to a unique level of aggregate consumption.

To gain some intuition about what is involved in this process, consider the case illustrated in Figure 3. Recall that this diagram represents the demand of an individual consumer. Suppose that, over successive periods, the price remains constant at  $p'$ . Consider any given period  $t$ , and let the quantity consumed by the individual in that period be  $q(t)$ . Let  $q'(t)$  be the level of consumption that is implied by the reference-dependent demand function that is defined with  $q(t)$  as the reference point; thus, actual consumption in each period is  $q(t) = q'(t - 1) + \varepsilon(q'[t - 1])$ . If  $q(t)$  is less than  $q_0$ ,  $p'$  is below the range of prices at which the relevant reference-dependent demand curve is vertical, and so  $q'(t)$  will be greater than  $q(t)$ . Conversely, if  $q(t)$  is greater than  $q_1$ ,  $p'$  is above the range of prices at which the relevant reference-dependent demand curve is vertical, and so  $q'(t)$  will be less than  $q(t)$ . If  $q(t)$  lies in the interval  $q_0 \leq q(t) \leq q_1$ , then  $q'(t) = q(t)$ , and so  $q(t + 1) = q(t) + \varepsilon(q[t])$ . Thus there is a tendency, independent of the noise in the model, for the quantity consumed to converge to the interval  $(q_0, q_1)$ ; within this interval, consumption drifts randomly as a result of noise.

Now consider the path of  $q(t)$ , starting from any arbitrary  $q(0)$ . As  $t$  increases, the expected value of  $q(t)$  will be less and less dependent on  $q(0)$ ; in the limit, as  $t$  tends to infinity, this expected value will be determined solely by the disturbance term  $\varepsilon(\cdot)$ . Let the expected value of  $q(t)$  in this limit be  $q^*$ . Then  $(p', q^*)$  is a point on the long-run demand function  $D^L$ .

<sup>19</sup> It would be simpler to make the value of the random variable independent of  $q'$ , but (since consumption cannot be negative) this would cause technical problems at very low values of  $q'$ .