Consumers' surplus when individuals lack integrated preferences: a development of some ideas from Dupuit

Robert Sugden

School of Economics, University of East Anglia

Norwich NR4 7TJ, United Kingdom

r.sugden@uea.ac.uk

25 June 2015

Abstract: In modern economics, consumers' surplus is understood as the sum of individuals' compensating variations, defined by reference to well-behaved preferences. If individuals lack integrated preferences, as behavioural economics suggests they often do, consumers' surplus cannot be defined. However, Dupuit – the earliest theorist of consumers' surplus – did not assume integrated preferences. His concept of consumers' surplus can be interpreted in terms of the maximum yield of discriminatory prices. In principle, this can be measured without making assumptions about preferences, but (contrary to what Dupuit apparently thought) is not in general equal to the area under the observed demand curve.

Keywords: consumers' surplus, price discrimination, integrated preferences, Dupuit

'Hence the saying which we shall often repeat because it is often forgotten: the only real utility is that which people are willing to pay for' (Jules Dupuit, 1844/ 1952, p. 262).

Consumers' surplus is one of the most important theoretical constructs in applied welfare economics. It plays an essential part in normative economic analyses of competition policy, of price regulation for natural monopolies, and of public provision of non-marketed goods such as road space, flood protection and free health care. It fits into a well-understood and internally consistent theoretical and philosophical framework. According to received economic theory, consumers' surplus is a measure of *compensating variation* – that is, of the change in an individual's money income which, if combined with the policy or project under consideration, would leave that individual's utility unchanged. In this theory, 'utility' is understood as an ordinal representation of an individual's stable preference ordering over relevant outcomes, and is assumed to be revealed in the individual's decisions. Philosophically, analyses that use consumer's surplus measurements are often interpreted as incremental exercises in *welfarism* – that is, exercises that aim to increase social welfare, where social welfare is understood as an aggregate of individuals' welfares, and preference-satisfaction is used as the criterion of individual welfare. But this whole framework rests on the assumption that individuals really do have stable and coherent preference orderings. What if they don't?

That is not an idle question. Research in experimental and behavioural economics has generated a large body of evidence of systematic patterns in individuals' choices that cannot plausibly be explained as the product of consistent preferences. Given what economic theory has traditionally regarded as a fully-defined decision problem, different 'frames' or 'cues' can activate different mental processes and generate different decision outcomes. For example, decisions may be influenced by apparently irrelevant but psychologically salient 'anchors' (e.g. Slovic and Lichtenstein, 1968; Ariely, Loewenstein and Prelec, 2003), by whether given changes in outcomes are framed as gains or losses (e.g. Kahneman and Tversky, 1979), and by whether the decision-maker's affective engagement with the problem is 'hot' or 'cold' (e.g. Loewenstein, 2005). Even under controlled laboratory conditions, individuals' responses to two presentations of exactly the same simple decision problem within the space of a few minutes can reveal a high degree of stochastic variation (e.g.

Becker, DeGroot and Marschak, 1964; Loomes, Moffatt and Sugden, 2002). Notice that the common feature of these cases is not that individuals are acting on preferences that violate conventional conditions such as transitivity, convexity or non-satiation. Nor is it that individuals are making errors in maximising given utility functions. It is that each individual's preferences, whatever their properties, are liable to change from one moment to another and from one choice problem to another. I shall say that such preferences are not *integrated*.

From the perspective of cognitive psychology, it is not particularly surprising that choice behaviour can be affected by cues that a theory of rational choice would treat as arbitrary, or that choice processes can be inherently stochastic. However, these observations pose severe problems for neoclassical welfare economics. The usual justification for using preference-satisfaction as a normative criterion is that each individual's preferences express her judgements about the determinants of her well-being. But many of the factors that have been found to influence preferences cannot plausibly be treated as relevant for well-being. If individual behaviour does not reveal integrated preferences, what are we to make of a form of normative analysis whose criterion is preference-satisfaction?

At first sight, it might seem that this problem can be safely ignored in many practical applications of the concept of consumer's surplus. In the context of competition and price regulation, or of the cost-benefit analysis of public projects, the consumers' surplus measurements that are usually needed are ones that aggregate over the behaviour of many individuals and over extended periods of time. If there is a stable demand function that allows average purchases of a good by consumers as a whole to be predicted from data about prices and average incomes, aggregate changes in consumer's surplus resulting from given price changes can be calculated using standard techniques. Does it then matter whether the individual consumers whose behaviour generated those data acted on integrated preferences? The premise of the present paper is that, *if we retain the received interpretation of consumers' surplus and of its role in normative economics*, this does matter.

According to the received interpretation, the object of welfare economics is to discover how best to secure individuals' well-being, and each individual's well-being is to be measured by the degree to which his assumedly integrated preferences are satisfied. Given these premises, data that aggregate across choices that (for any given individual) have been influenced by different arbitrary cues at different moments are not fit for purpose. That is true even if, at the level of aggregation we are using, the data *could have been* generated by a

hypothetical set of ideally rational consumers: the object is supposed to be to satisfy the preferences of actual consumers, not those of imaginary ones.

But do we have to accept the interpretation of consumers' surplus that leads to this negative conclusion? I will argue that an alternative interpretation is possible – one which breaks the theoretical link between consumers' surplus and integrated preferences. The intuition behind my proposal is expressed in the quotation from Dupuit that I have used as my epigraph. Dupuit is justly recognised as the founding father of cost-benefit analysis and as the first economist to develop a concept of consumers' surplus. For Dupuit, this concept is based on what he treats as a fundamental principle of economics: *the only real utility is that which people are willing to pay for*. I will ask whether it is possible to define a concept of consumers' surplus that is based on what individuals are willing to pay for a good, at the moment at which they buy it, and in the frame in which they buy it. Such a concept would not need any concept of an individual's preferences, utility or well-being that extended across different moments or different frames. I will argue that Dupuit's analysis of consumers' surplus suggests that he had some such concept in mind.

1. Behavioural welfare economics and the opportunity criterion

My proposed interpretation of consumers' surplus fits into an approach to the problem of reconciling normative and behavioural economics that I have developed and defended in previous papers (Sugden, 2004, 2007, 2013; McQuillin and Sugden, 2012). In this section, I give a brief account of this approach, against the background of a more orthodox alternative.

A consensus seems to be developing among behavioural economists in favour of a particular strategy for dealing with normative issues, which I will call *behavioural welfare economics*.[1] This strategy, proposed by (among others) Sunstein and Thaler (2003), Bernheim and Rangel (2007) and Kőszegi and Rabin (2007), uses the satisfaction of revealed preferences as the normative criterion whenever those preferences are well-defined and free of 'mistakes'. However, when an individual's choices between options vary according to (supposedly) arbitrary context-specific cues, the preferences revealed in those choices are not necessarily respected. Instead, the welfare economist tries to reconstruct the preferences that the individual would have revealed, had she not made mistakes; implicitly, these *latent*

---

[1] For a fuller account of this approach, see the review and critique by Infante, Lecouteux and Sugden (2015).

preferences are assumed to be context-independent. In Bernheim and Rangel's variant of this approach, preferences are guaranteed respect only if they are revealed in 'non-suspect' situations in which the individual's reasoning is not likely to be impaired. Clearly, the logic of behavioural welfare economics implies that consumer's surplus measurements are not suitable for use in applied welfare economics unless one can be confident that they are based on context-independent preferences that are error-free (or non-suspect).

In my proposed approach, in contrast, the normative criterion is not well-being or the satisfaction of integrated preferences; it is *opportunity*. The essential idea is that an individual's opportunities are described by the set of alternative time-profiles of consumption and holdings of goods that he can attain by acts of choice. This set is the individual's *opportunity set*. Unambiguous expansions of an opportunity set are treated as improvements for the individual concerned, because they allow him greater scope to act according to his preferences – whatever these may turn out to be and whether or not they are stable from one decision problem to another. If one presupposed the normative framework of mainstream welfare economics, this criterion would perhaps be hard to defend. But the framework I use makes two significant breaks with mainstream welfare economics, which together make the criterion less counter-intuitive.

The first break is with the idea that normative economics is addressed to an imagined 'social planner' whose objective is to maximise the overall welfare of a society, and that the object of normative analysis is to arrive at 'policy implications' that this planner would want to implement. My approach is *contractarian* in the sense of Buchanan (1968, 1975). By this, I mean that normative economics is addressed to private individuals, viewed as potential parties to voluntary agreements. Given this approach, the relevant question to ask about a normative criterion such as preference-satisfaction or opportunity is not whether it corresponds with the relevant individual's well-being, as that is viewed by a benevolent social planner. It is whether it represents that individual's interests, *as he or she perceives them*.

The second break is with the idea that the continuing identity of a person across decision problems is represented by the existence of an integrated set of consistent preferences, and hence that an individual who acts on non-integrated preferences is not a unitary decision-making agent. There is a long tradition in economics of modelling a person who acts on inconsistent preferences as a collection of distinct *transient selves*, each of which is then treated as an agent with its own preferences, interacting strategically with other selves.

This modelling strategy is often used to represent problems of time-inconsistency and self-control (e.g. Strotz, 1955–56; Schelling, 1984). Normative analysis then usually proceeds either by privileging the preferences of one of the selves (typically the self that is judged most rational or reflective) or by treating the different selves as though they were distinct individuals. My approach differs by treating the identity of a person over time as a continuing *locus of responsibility*. To understand your identity in this way is to accept the authority of each of your transient selves with respect to the decisions that it is called on to make, rather than to evaluate all those decisions against a single integrated set of preferences. This allows you to see any expansion in your opportunity set as good for you as a continuing person, without needing to ask what your future preferences will be.

In this framework, the analogue of Pareto efficiency as a normative principle is the *Opportunity Criterion*. This criterion assesses *regimes* for a given economy, a regime being a profile of opportunity sets, one such set for each individual. Intuitively, the Opportunity Criterion requires that all opportunities for feasible transactions that individuals might want to make are available to them in their respective opportunity sets. 'Might want' here is to be interpreted without reference to individuals' *actual* preferences. Rather than treating preferences as given, the criterion takes the viewpoint of individuals who have not yet thought about, not yet decided, or not yet discovered what their preferences will be. Thus, it takes account of the whole range of preferences on which each individual *might* want to act – including preferences that are context-dependent or that vary from moment to moment during the process of trading. The only assumption that is made about preferences (and this is made only implicitly) is that in the case of 'money' – the good that serves as the medium of exchange – more is always preferred to less. (Notice that this is not an assumption about the consistency of an individual's preferences across contexts or across time. It is an assumption about behaviour that applies in any given context and at any given time.) Let us say that a person is *willing to pay for* something if (at the moment at which he chooses to buy) he is willing to pay enough to make others willing to play their parts in supplying it to him (at the moments at which they choose to sell). A regime that satisfies the Opportunity Criterion, one can then say, allows each individual to get whatever he wants and is willing to pay for, whenever he wants it and is willing to pay for it. One can also say that, if the Opportunity Criterion is satisfied, there is no feasible putative transaction such that everyone who would be party to it wants to engage in it, but it has not been realised. Thus, that criterion can be interpreted as equivalent to the absence of unrealised opportunities for profitable arbitrage. It

can be shown that the Opportunity Criterion is satisfied in any exchange economy in which the single-price law of markets holds and all markets clear, irrespective of whether individuals act on integrated preferences (Sugden, 2004; McQuillin and Sugden, 2012).

The Opportunity Criterion, as so far formulated, is adapted for the analysis of general equilibrium in exchange economies. It does not have direct application to the partial equilibrium problems of cost-benefit analysis that Dupuit was concerned with and for which consumers' surplus measurements are most useful. However, I believe that the normative intuitions that underlie the Opportunity Criterion have some bearing on such problems.

Consider Dupuit's paradigm example of a problem for cost-benefit analysis – the footbridge. Suppose that an entrepreneur is considering building a footbridge over a river and charging tolls for crossing it. He will incur costs in building and maintaining the bridge, but these costs are independent of the volume of traffic it will carry. Even if no tolls were charged, the bridge would have more than enough capacity to meet the demand for it. If the entrepreneur can design a tariff that will generate revenue in excess of costs, he will build the bridge; if not, not. If the bridge is built and is profitable, one might say that the entrepreneur has mediated a mutually beneficial transaction between the bridge users and those people who supply the inputs necessary for building and operating the bridge.

In the spirit of the Opportunity Criterion, one might propose a criterion requiring that all opportunities for this kind of profitable arbitrage are exhausted, and that the actual profits made by arbitrage are zero. The activity of bridge-building would satisfy such a criterion if the following conditions held: (1) price discrimination is capable of appropriating any desired proportion (up to the whole) of consumers' willingness-to-pay; (2) bridges are built if and only if they are capable of being financed by discriminatory pricing; (3) the revenue generated by the prices actually charged exactly covers total costs; and (4) those prices do not deter any user who is willing to pay the marginal costs he imposes. Clearly, these are not conditions that one can expect to be satisfied as an unintended outcome of profit-seeking behaviour in an unregulated market. But one might think of them as an ideal to which the regulation of markets should aspire. Designing a regulatory regime that approximates as closely as possible to this ideal is an example of the type of problem in which consumers' surplus measurements are typically used. This is precisely the problem that Dupuit set out to solve, and for which he developed the concept of consumers' surplus.

2. Dupuit's analysis of consumer's surplus

The founding text of cost-benefit analysis is Dupuit's (1844/ 1952) paper, 'On the measurement of the utility of public works'. This paper was intended to form a chapter in a never-completed book about the application of political economy to public works. The primary subject of the paper is 'the conditions which [public] works must fulfil in order to be really useful' or 'How, in a word, is public utility to be measured?' (p. 255); Dupuit's examples of public works include roads, railways, canals, bridges and water supply systems. In trying to answer this (as it might seem) practical question, he engages with debates, central to mid-nineteenth century economics, about the nature of utility and value.

Many features of Dupuit's analysis will strike a modern economist as far ahead of its time. What are now seen as fundamental principles of cost-benefit analysis – not only consumers' surplus, but also the distinctions between real resource costs and transfer payments and between pecuniary and technological externalities, are presented for the first time. Dupuit also presents the first analysis of price discrimination. But there is one respect in which his analysis is strikingly different from later neoclassical theory: in theorising about choices made by consumers, he makes no explicit assumptions about rationality or utility-maximisation. By this, I do not mean to imply that Dupuit had no inkling of, or rejected as invariably false, the hypothesis that consumers seek to maximise the satisfaction they derive from the goods they are able to buy. Historians of economic thought have pored over the economic writings of the mid-nineteenth century looking for insights that prefigure the 'marginal revolution' of the 1870s. It is not my intention to add to this enterprise. My concern is only with how Dupuit uses the concept of consumers' surplus in his 1844 paper. For my purposes, what matters is that he formulates this concept without making explicit assumptions about consumer rationality.

Indeed, Dupuit explicitly rejects the idea that the methods of political economy can be used to explain or predict consumer choice. For example: 'The variable, yea mobile, nature of the value of utility is indeed well known to business men and has long been exploited by them. That is what lies behind all transactions that are sheltered from competition' (p. 260). This remark about the behaviour of businesses that are sheltered from competition is a preface to Dupuit's analysis of price discrimination. To a modern reader, his account of consumer choice is more behavioural than neoclassical:

> The same commodity in various guises is often sold in different shops at quite
> different prices to the rich, the moderately well-off, and the poor. The fine, the
> very fine, the superfine, and the extra fine, although drawn from the same barrel
> and although alike in all real respects other than the superlative on the label, sell at

widely different prices. Why? Because the same thing has a widely differing utility depending on the consumer. If there were only one medium price, there would be a loss to those who did without the product because its utility to them was less than that price, and a loss to the seller who, from many buyers, would be receiving payment for only a fraction of the utility of the services rendered. (p. 261)

Dupuit is clearly assuming that consumer choice (at least, when consumers are comfortably-off) is governed by preferences that are labile and context-dependent, and that can be exploited by profit-seeking suppliers.

After defining what would now be called a demand function as a series of (price, quantity) pairs or 'relationships', he writes:

This series of relationships is not known for any commodity, and it can even be said that it will never be known since it depends on the volatile will of human beings; it is today no longer what it was yesterday. It is thus of no avail to try to determine this relationship exactly by experience or groping experiment, but there do exist certain general laws to which the relationship, in its very volatility, remains constantly subject. (p. 277).

Significantly, these 'general laws' are properties of functions which aggregate the demands of many individuals, and are justified as 'fact[s] of experience' which have been verified statistically, rather than as principles that can be explained in terms of rational choice by individuals. Dupuit states two such laws. The first is that demand curves slope downwards. The second is that, for any given demand curve, the gradient is more negative at higher prices. This second law is explained as reflecting the 'pyramid' structure of social classes in the population over which demand functions aggregate (p. 277). Implicitly, Dupuit also treats as a law the regularity that demand curves are approximately continuous, a regularity that is naturally interpreted as a result of aggregation across individuals.

Since Dupuit's method of measuring the utility that consumers derive from public works uses only the data in the aggregated demand function, the implication is he believes that this measure is valid even if (as he clearly thinks is often the case) consumers' preferences are 'volatile'. In other words, he is presenting consumers' surplus as a measure that does not depend on assumptions about the integrated preferences of individual consumers. If he is right about the validity of his measure, the fact that it can be calculated from observable (or predictable) data about the behaviour of consumers in the aggregate is a huge merit – a merit that would be particularly salient to an engineer thinking about the practicalities of assessing the costs and benefits of public works.

Despite treating utility as volatile, Dupuit maintains that the concept has a well-defined meaning: *utility is whatever people are willing to pay for*. Thus, the value of any item of consumption to the consumer is the surplus of what the consumer is willing to pay over what he actually pays: 'We see that in general the relative or definitive utility of a product is expressed by the difference between the sacrifice which the purchaser would be willing to make in order to get it, and the purchase price he has to pay in exchange' (pp. 262–263). Dupuit makes clear that this definition of 'utility' is designed for the purposes of political economy:

> We say political economy, because this is not, in the last analysis, a rigorous measure of the quality which things have *of being able to satisfy men's needs* [a reference to Say's definition of utility]; it would be difficult to say whose hunger was the greater – the rich man's, who would be willing to give a million for a kilogram of bread, or the poor man's, who, having nothing else to give, would risk his life for it. But political economy, being concerned only with wealth, can take account of the intensity of a wish only through its monetary expression. Political economy only bakes bread for those who can buy it, and leaves to social economy the care of supplying it to those with nothing of value to give in exchange. (p. 262).

Notice that Dupuit considers and rejects the idea of measuring the intensity of individuals' desires for consumption or the intensity of the satisfaction they derive from it: his objective is simply to measure *willingness to pay*.

Dupuit presents consumers' surplus (in his terminology, consumers' 'relative utility') as a measurement tool that can be useful in providing guidance to legislators in making decisions about the provision and pricing of public works (pp. 261, 279). Repeatedly, he describes consumers' surplus as the sum, over all units actually consumed, of the maximum additional tax or toll that could be levied on each unit without preventing it from being consumed. Thus, to calculate the 'utility of transportation' of freight on a road system, he imagines 'a traffic tax increasing little by little, which would cause the successive disappearance of several of the tons together comprising the 50 million tons being carried on these roads'; each ton is then multiplied by 'the tax which would prevent it from moving' (p. 271). The same principle applies to passenger transport. He considers the case of a new railway which diverts traffic from a lower-priced but slower stagecoach service. The utility gain for rail users is 'the toll [on railway travel] which would dissuade passengers from stepping out of the stagecoach and into the railway carriage' (p. 274). It is significant that Dupuit uses the same method of utility measurement for freight transport (much of which, presumably, is an intermediate good bought by manufacturers and retailers) and for passenger

transport (which is often a final consumption good). This is possible because his concept of willingness to pay is not tied to subjective desires or satisfactions: it is simply the highest price that a potential buyer would in fact pay.

One way of describing Dupuit's approach is to say that the sum of consumers' actual payments and consumers' surplus is the maximum amount of money that, in principle, could be raised from the sale of the quantity actually consumed using 'perfect' price discrimination. The problem with this statement, of course, is the need to spell out the meaning of 'perfect'. In modern neoclassical economics, the issue is seen the other way round. Starting from assumptions about well-behaved preferences and utility-maximisation, the neoclassical approach defines consumers' surplus as a sum of compensating variations; perfect price discrimination is then defined as pricing that appropriates the whole of this potential surplus. But that analysis depends on the assumption that each consumer acts on preferences that are integrated and well-behaved. In contrast, Dupuit seems to want to claim that he has a well-defined concept of perfect price discrimination, and a method of measuring the revenue it would yield, which does not require that assumption.

Dupuit's understanding of the link between consumers' surplus and price discrimination underlies his advocacy of price discrimination as a means of recovering fixed costs. Thus, although the practices of price discrimination used in business work by setting 'traps for the buyer's vanity and his credulity', he thinks they are often 'more equitable and fairer than one might expect at first sight' and might be good examples to be followed in setting tariffs for public works (p. 261). Anticipating the content of subsequent (but never written) chapters of his book, Dupuit says that he hopes to show how such tariffs can be fixed 'according to rational principles, in order to produce the greatest possible utility and at the same time a revenue sufficient to cover the cost of upkeep and interest on capital' (p. 271). The implication is that a rational tariff would use price discrimination to recover total costs while, as far as possible, not deterring any user who was willing to pay the marginal cost of supply.

Recall Dupuit's claim that his concept of utility – that utility is whatever people are willing to pay for – is appropriate for political economy because the subject matter of political economy is wealth. The implication seems to be that the production and sale of a good is wealth-creating if and only if consumers' willingness to pay for it exceeds the cost of supply. Thus, if the costs of a proposed public project can be more than recovered by discriminatory pricing, that is evidence that the project would create wealth.

This idea emerges clearly in Dupuit's example of the footbridge. In this case, the 'absolute utility to society' of the bridge is the total of consumers' willingness to pay; the 'relative utility' is this total less the costs of maintenance and interest on capital (p. 270). Here, by implication, Dupuit means relative utility *to society*, as contrasted with relative utility *to consumers* (which, as he explains in the passage I quoted earlier, is the excess of consumers' willingness to pay for a good over what they have to spend to buy it). If the two concepts of relative utility are not equal, the difference is the net surplus or profit accruing to producers of the good. Thus, 'relative utility to society' corresponds with what, in modern cost-benefit analysis, would be called 'net social benefit' – the sum of net benefits to consumers and producers.

In his analysis of (what are now called) pecuniary externalities, Dupuit allows costs and benefits to different individuals, if equal and opposite in money value, to cancel out in the measurement of public utility. His leading example is concerned with the measurement of the benefits of constructing a canal that will carry coal into an area with no local coal deposits. The opening of the canal causes a fall in the local price of coal. This benefit of the canal is measured by the maximum revenue that in principle could be generated by additional taxes on coal traffic on the canal. But suppose that wood is a locally-produced substitute for coal. Although wood is not carried on the canal, the new supply of coal leads to a fall in the local price of wood. Should the increase in consumers' surplus on wood count as an additional benefit of the canal? In the earliest statement of what is now recognised as a fundamental principle of cost-benefit analysis (see, e.g., Sugden and Williams, 1978, pp. 134–147), Dupuit argues that the answer is 'No', because when a project induces indirect effects on market prices, 'the consumer gains what the producer loses, or *vice versa*'. These offsetting gains and losses are 'merely changes in the distribution of wealth', and so are not relevant for 'the calculation of utility', even though they are matters of concern for the state (p. 272).

For Dupuit, then, utility measurement is fundamentally concerned with the amounts of money that people are willing to pay for goods, and this willingness to pay is interpreted in terms of the maximum revenue that can in principle be generated by discriminatory pricing. However, his method of actually making these measurements uses only the data that is contained in observable demand functions. He starts with a demand curve (he calls it a 'curve of consumption') which plots quantities of some good consumed by a population of consumers as a function of the price. I will write this function as $x = f(p)$, where x is quantity

12

consumed and *p* is price. Most of Dupuit's analysis works with the inverse function $p = f^{-1}(x)$. He interprets 'quantity' as numbers of 'articles', ranked in inverse order of the highest price at which they are bought. (That is, article 1 is the article that is bought when the price is so high that only one article is bought, article 2 is the additional article that is bought at the slightly lower price at which exactly two articles are bought, and so on.) Thus, a value $f^{-1}(x)$ is defined for each article. In modern terminology, this is the marginal valuation of the good, or consumers' marginal willingness to pay for it, evaluated at *x*. Dupuit calls it the 'utility' of the specific article. Given any price *p*, the total utility of the quantity *f(p)* is defined as the integral of $f^{-1}(x)$ between 0 and *f(p)* (that is, as the sum of the utilities of the articles that are consumed). If consumers' actual expenditure *pf(p)* is subtracted from this integral, we arrive at Dupuit's formula for consumers' surplus. In deriving this formula, Dupuit makes no explicit assumptions about consumers' behaviour beyond the observations that are described by the demand curve itself. In particular, he makes no explicit assumptions about preferences or utility-maximisation. However, the formula itself is essentially the same as that used in modern exercises in practical cost-benefit analysis and in modern definitions of perfect price discrimination. (The main difference is that Dupuit treats the area under an observed demand curve as an *exact* measure of surplus, rather than as a close approximation that may need to be corrected to take account of income effects.)

Dupuit claims to have developed a coherent concept of consumers' surplus that does not depend on assumptions about integrated preferences. Viewed in the light of the findings of behavioural economics, this claim is of great interest. However, it has not received much attention. It has been seen as unimportant because, in the dominant neoclassical tradition of economics, the assumption of integrated preferences has been treated as unproblematic. Thus, those features of Dupuit's analysis that allow him to avoid making assumptions about integrated preferences – particularly his definition of utility in terms of willingness to pay – have been seen as obsolete or even confused.

3. Dupuit, Walras and the neoclassical concept of consumers' surplus

As Dupuit's treatment of utility is so different from that of neoclassical economics, it is perhaps not surprising that it was summarily dismissed by Walras (1900/ 1954) as an 'egregious error' (p. 443). Walras recognises that, for Dupuit, the 'measure of utility' is 'the maximum pecuniary sacrifice which a consumer is willing to make in order to procure a unit of product', and he explains how Dupuit measures total utility by 'aggregating layers of gross

receipts' at different prices. Walras's criticism seems not to be directed at Dupuit's method of *measuring* maximum pecuniary sacrifice, but rather at the claim that this is a measure *of utility*:

> To be sure, the maximum pecuniary sacrifice which a consumer is willing to make in order to obtain a bottle of wine, for instance, depends in part on the utility of this bottle of wine for the consumer… . [But in general] the maximum pecuniary sacrifice which a consumer is willing to make in order to obtain a unit of a product depends not only on the utility of the product in question, but also on the utility of all the other products in the market, and, finally, on the consumer's means. (p. 445).

Thus:

> We may, therefore, definitely reject all Dupuit's statements in his two memoirs which bear upon the variation of utility as price varies and as the quantity demanded varies from price to price. [These statements] rest on a confusion of ideas resulting from Dupuit's complete failure to distinguish between utility or want curves on the one hand, and demand curves on the other. (p. 446)

There seems to be general agreement among historians of economic thought that Walras was ungenerous in not recognising Dupuit's achievement in developing the first theoretical analysis of consumers' surplus. Nevertheless, Walras's criticisms are usually seen as broadly correct. For example, Ekelund and Hébert (1985: 422–423) say that Dupuit failed to distinguish between marginal utility curves and demand curves, with the result that his consumers' surplus measure 'tends to misstate true utility'. Tubaro (2006: 11) says that Dupuit's writings are 'somewhat confused about the distinction between utility and demand'. Houghton (1958) is more ambivalent, saying that 'Dupuit's implied confusion (identification?) of demand and utility curves' was 'a much less serious blunder (abstraction?) than Walras believed'. (I take it that Houghton does not want to assert that Dupuit's analysis was confused or erroneous, but still sees the absence of a distinction between utility and demand as a weakness that later theorists have overcome.)

The consensus account of the evolution of the theory of consumers' surplus is encapsulated in Ekelund and Hébert's (1985: 445) metaphor of an upward climb of progress with three 'plateaus' of achievement, successively reached by Dupuit, Marshall (1900) and Hicks (1941). Marshall's advance on Dupuit was to distinguish between utility and demand and to derive the conditions under which Dupuit's measure of surplus is a correct money measure of the utility that an individual derives from a good – namely that the individual's tastes, the individual's marginal utility of income, and the prices of other goods all remain

constant.[2] Hicks's advance on Marshall was to use ordinal rather than cardinal utility and to derive Marshall's concept of surplus as a special case of the concept of compensating variation.

Notice, however, that the advances made by Marshall and Hicks work by assimilating Dupuit's analysis to a theoretical framework in which consumers' choices are rational, rationality being modelled as the maximisation of utility (or, equivalently, as choosing what is ranked highest in a preference ordering that can be represented by a utility function). In this framework, utility is a more fundamental concept than demand, because demand functions can be derived from utility functions and budget constraints by solving constrained maximisation problems. Marshall and Hicks are able to state and prove theorems that cannot be expressed in Dupuit's framework. But how far these theorems are useful advances depends on how successfully the hypothesis of utility maximisation explains consumers' actual choices. If in fact those choices reveal integrated and neoclassically well-behaved preferences, Dupuit's analysis might reasonably be said to have been superseded by later developments. But if they do not, Dupuit's concept of utility as whatever people are willing to pay for might be more useful than the neoclassical concepts that replaced it.

4. Was Dupuit right? The corner-shop model

In this section, I consider whether Dupuit was justified in claiming that his measure of consumers' surplus is valid even if consumer behaviour is volatile. I do this by developing a model which displays three different concepts of the 'absolute utility' that an individual consumer derives from consuming a given quantity of a good. The first concept is *Dupuit's measure* – the area under the observed demand curve. The second concept is *willingness to pay*, interpreted as the maximum revenue that can in principle be appropriated from the sale of the relevant quantity of the good by pre-announced discriminatory pricing. Dupuit effectively treats these two concepts as equivalent. The third concept is the *experienced utility* – utility in the classical utilitarian sense of hedonic experience – enjoyed by the consumer. Experienced utility is a close relative of a concept that Dupuit considers but judges inappropriate for political economy – the degree to which the consumer's needs are satisfied. My object is to investigate the relationships between these three concepts when the

---

[2] Houghton (1985) argues convincingly that these developments by Marshall were anticipated by Auspitz and Lieben (1889).

consumer does not act on integrated preferences. In particular, I am interested in whether Dupuit's measure is equivalent to willingness to pay.

The name I have given my model alludes to an era when, for most people living in towns, there were small grocery shops (often on street corners) within a few minutes' walk of their homes. This made it possible for people to delay buying items of household consumption until the moment they were actually needed. The model is of one consumer's decisions over three *periods*, $t = 1, 2, 3$ (or morning, afternoon and evening), of a single day. There are two goods, *money* and *food*. At the start of the first period, the consumer is endowed with money but no food. Food comes in discrete units. In each period t, the consumer must consume either zero units of food (denoted by $x_t = 0$) or one unit ($x_t = 1$). There is a *tariff* which specifies, for each period, the terms on which money can be exchanged for food. Food that is bought in one period can either be consumed in that period or costlessly stored for later consumption. At the prices that will be considered in the model, the consumer always has more than enough money to buy as much food as she can use, but she also derives utility from money that is not spent on food. (As a convenient and harmless simplification, I assume that the consumer is not permitted to buy food that she does not consume, even if the price is zero.) The problem is to define and measure the 'absolute utility' that the consumer would derive if the price of food were zero in every period.

I assume that in period 1, the consumer's experienced utility is $g_1(x_1) - o_1$, where o_1 is her *outlay* of money for the purchase of food in that period; $g_1(x_1)$ is her *gross utility* in that period. In period 2, her experienced utility is $g_2(x_1, x_2) - o_2$; in period 3 it is $g_3(x_1, x_2, x_3) - o_3$. This specification allows the gross utility derived from consumption in one period to depend on previous consumption. (For example, the individual may get more utility from eating, the hungrier she is at the time.) I assume that, in each period t, gross utility is greater (other things being equal) if $x_t = 1$ than if $x_t = 0$. I normalise by setting $g_1(0) + g_2(0, 0) + g_3(0, 0, 0) = 0$. Notice that this specification implicitly assumes that the marginal utility of money is constant, and measures utility in money units. This abstracts from the complications that would be generated by income effects. I define A^E, the experienced-utility measure of the gross (or 'absolute') utility derived from consuming in all three periods, as $g_1(1) + g_2(1, 1) + g_3(1, 1, 1)$.

I define a *demand function* in terms of the consumer's total purchases of food over the day, conditional on the tariff having a *unitary price* structure – that is, there is a single price that applies to all units of food purchased in all periods. This function is written as $X = f(p)$ where X is the total quantity of food bought in the three periods combined and p is the unit price. This function is fully described by the triple (h_1, h_2, h_3) where h_1, h_2 and h_3 are respectively the highest prices at which the consumer would buy 1, 2 and 3 units. Thus A^D, Dupuit's measure of the absolute utility derived from consuming in all three periods, is $h_1 + h_2 + h_3$.

As a starting point, I consider the *neoclassical* case in which the consumer acts on the decision rule of maximising the sum of experienced utility over the three periods – that is, maximising $g_1(x_1) + g_2(x_1, x_2) + g_3(x_1, x_2, x_3) - (o_1 + o_2 + o_3)$. It is convenient to define $G(X)$ as the maximum value of $g_1(x_1) + g_2(x_1, x_2) + g_3(x_1, x_2, x_3)$ that can be achieved, given that the quantities purchased in the three periods sum to X. Thus, since no more than three units of the good can be bought, the experienced-utility measure of absolute utility is given by $A^E = G(3)$.

Now consider the consumer's demand function. Since this is defined with reference to a unitary price structure, the decision rule from which this function is derived can be expressed as the maximisation of $G(X) - pX$. In this formulation, the problem of optimising the distribution of consumption over the day is hidden from view; the consumer is represented as choosing consumption *for the day*, given the unit price *for the day*. This is an entirely standard neoclassical decision problem in which the consumer acts on integrated preferences. Since income effects have been assumed to be zero, it follows from well-known results about consumers' surplus that Dupuit's measure A^D of the area under the demand curve is equal to the experienced-utility measure A^E.

Here is one way of explaining why this is so. Consider any two prices p and $p - \Delta p$, where Δp is a very small positive number. Let X and $X + \Delta X$ respectively denote the consumer's chosen consumption at prices p and $p - \Delta p$. The increase in her experienced utility as a result of the fall in the price must be at least $X\Delta p$, because she has the option of consuming X at the lower price. But this increase cannot be greater than $(X + \Delta X)\Delta p$, because she has the option of consuming $X + \Delta X$ at the higher price. Now take some price p^* at which chosen consumption is zero. Clearly, experienced utility is zero at this price. Consider the effect of reducing the price from p^* to zero in a sequence of very small changes. It

follows from the argument I have just given that the total increase in experienced utility is approximately equal to the area under the demand curve, and that the two measures approach one another as the price increments used in the calculation approach zero. Thus, $A^{\mathrm{D}} = A^{\mathrm{E}}$. Notice that this argument depends not only on the assumed absence of income effects and on the measurement of utility in money units, but also (and crucially) on the assumption that the consumer maximises experienced utility.

What about the willingness-to-pay measure of absolute utility? Recall that this measure, which I will write as A^{W}, is the maximum revenue that can be appropriated by pre-announced discriminatory pricing. Given the assumption that the consumer maximises experienced utility, it is immediately obvious that $A^{\mathrm{W}} = A^{\mathrm{E}}$. Since the consumer always has the option of not consuming and thereby getting zero experienced utility, she will not consent to pay more than A^{E} to consume three units. But if she faces a tariff with a fixed charge of A^{E} and a marginal charge of zero per unit consumed, she will be indifferent between accepting and rejecting the offer. Thus, in the neoclassical case, $A^{\mathrm{W}} = A^{\mathrm{E}} = A^{\mathrm{D}}$.

I now consider a different possibility. This is that the consumer's decision rule is *myopic*: in each period, she maximises experienced utility *in that period*. In general (and as will emerge from my analysis), this rule does not generate choices that reveal integrated preferences. I will investigate the implications of this decision rule in combination with three alternative assumptions about experienced utility.

In *Specification 1*, experienced utility is *separable* in the following sense: in each period, gross utility depends only on consumption in that period. In other words, the value of $g_2(x_1, x_2)$ is independent of x_1 and the value of $g_3(x_1, x_2, x_3)$ is independent of x_1 and x_2. I also stipulate that gross utility is zero in any period t in which $x_t = 0$. It is easy to see that, under this specification, the demand function for a myopic consumer is exactly the same as it would be for a neoclassical consumer.[3] Thus, as in the neoclassical case, $A^{\mathrm{D}} = A^{\mathrm{E}}$. It is also easy to see that A^{E} can be appropriated by a tariff in which, in each period, the unit price is equal to the gross utility of one unit of consumption in that period, and that no more than this can be appropriated. Thus, again as in the neoclassical case, $A^{\mathrm{W}} = A^{\mathrm{E}}$. So $A^{\mathrm{D}} = A^{\mathrm{W}}$: Dupuit's measure of absolute utility is equivalent to willingness to pay.

---

[3] Recall that the demand function is defined with respect to a unitary price structure. In such a price structure, there is no advantage in buying a unit of food before the period in which it is consumed. If unit prices differed between periods, a myopic consumer would not take up advantageous opportunities for buying in advance of consumption, but a neoclassical consumer would.

In *Specification 2*, experienced utility is characterised by *diminishing satisfaction*. The experienced utility of any unit of food is fully determined by whether it is the first, second or third unit consumed in the day (irrespective of the period in which it is consumed); the first unit gives most utility and the third unit gives the least. Thus $g_1(1) = g_2(0, 1) = g_3(0, 0, 1) > g_2(1, 1) = g_3(0, 1, 1) = g_3(1, 0, 1) > g_3(1, 1, 1)$. Again, I stipulate that gross utility is zero in any period t in which $x_t = 0$. As in Specification 1, the demand function for a myopic consumer is the same as it would be for a neoclassical consumer; this function is defined by values of h_1, h_2 and h_3 that are respectively equal to the experienced utilities of the first, second and third units consumed in the day. Thus $A^D = A^E$. The maximum revenue that can be appropriated is A^E; this can be appropriated by a tariff which charges h_1 for the first unit consumed in the day, h_2 for the second and h_3 for the third. Thus, as in Specification 1, $A^W = A^E$. Again, Dupuit's measure of absolute utility is equivalent to willingness to pay.

In *Specification 3*, experienced utility is characterised by *increasing deprivation*. The gross utility of consuming a unit of food is independent of previous consumption and is the same in all periods. However, the gross utility of *not* consuming food in any period is determined by the number of periods that have elapsed since the last consumption of food (or since the start of the day if there has been no previous consumption). The greater the number of such periods, the less is the gross utility. To simplify the exposition, and because my purpose is to exhibit a counter-example to the supposition that $A^W = A^E = A^D$, I will use a numerical example. In this example, the gross utility of consuming a unit of food in any period is 2. The gross utility of not consuming is 1 if no periods have elapsed since the last period of consumption; it is 0 if one such period has elapsed; and it is –1 if two such periods have elapsed. Thus (using * to denote that the relevant variable can take either of the values 0 or 1):

$$g_1(1) = g_2(*, 1) = g_3(*, *, 1) = 2;$$
$$g_1(0) = g_2(1, 0) = g_3(*, 1, 0) = 1;$$
$$g_2(0, 0) = g_3(1, 0, 0) = 0;$$
$$g_3(0, 0, 0) = -1.$$

Notice that the normalisation $g_1(0) + g_2(0, 0) + g_3(0, 0, 0) = 0$ is preserved.

As a benchmark, it is useful to consider the demand function of a neoclassical consumer, given this specification of experienced utility. If a utility-maximising consumer chooses not to buy any food, her total gross utility will be zero. If she chooses to buy exactly

one unit of food, she will consume it in period 2, getting a total of 4 units of gross utility. If she chooses to buy exactly two units, she will be indifferent about when she consumes them, and will get a total gross utility of 5. If she buys three units, her total gross utility will be 6. From the preceding propositions, it is easy to work out that the demand function is given by $h_1 = 4$, $h_2 = 1$, $h_3 = 1$. Since we are dealing with a neoclassical consumer, we can conclude that $A^W = A^E = A^D = 6$.

But now consider the myopic consumer. As a first step, I derive her demand function. To avoid unnecessary complications, I assume that if buying and not buying give the same utility, the consumer buys. If $p \leq 1$, the consumer will buy one unit in each period. If $1 < p \leq 2$, she will not buy in period 1, but she will buy in period 2. Having consumed in period 2, she will not buy in period 3. If $2 < p \leq 3$, she will not buy in periods 1 and 2, but she will buy in period 3. If $p > 3$, she will not buy at all. Thus, the demand function is described by $h_1 = 3$, $h_2 = 1$, $h_3 = 1$. So A^D, Dupuit's measure of absolute utility as the area under the demand curve, is 5. This is despite the fact that A^E, the experienced-utility measure of absolute utility, is 6.

What about A^W, the willingness-to-pay measure of absolute utility? It turns out that no tariff can appropriate more than 3. Suppose that the consumer is induced to consume in all three periods. To be induced to consume in period 1, she must be able to buy one unit in period 1 at a price no greater than 1. Given that she has consumed in period 1, if she is to be induced to consume in period 2, she must be able to buy one unit in period 2 at a price no greater than 1. And similarly for period 3. So the maximum surplus that can be appropriated is 3. Similar reasoning establishes that the maximum surplus that can be appropriated is also 3 if the consumer is induced to consume exactly two units, or if she is induced to consume exactly one unit.

The implication is that if consumers act on non-integrated preferences, Dupuit's measure of absolute utility using observable demand data does not always correspond with the potential yield of perfectly discriminatory pricing. I have conjectured that Dupuit thought that the two measures necessarily coincide. If this conjecture is correct, one possibility is that, despite his remarks about the volatile will of human beings, Dupuit assumed that human beings act on integrated preferences. Another possibility, which is perhaps more consistent with the text, is that he had in mind some less restrictive assumption about behaviour which

would be sufficient to imply the equivalence of the two measures. My Specifications 1 and 2 are examples of such assumptions. Or, of course, he could simply have been mistaken.

5. Conclusion

From my point of view, as an economic theorist rather than an historian of thought, it is more important to focus on Dupuit's guiding principle, that the only real utility is that which people are willing to pay for. I have argued that this principle is best understood as implying that the gross benefit (or 'absolute utility') resulting from the consumption of a given quantity of a good is the maximum revenue that could be raised from the sale of that quantity at perfectly discriminating pre-announced prices. Dupuit's principle coheres with an approach to normative economics which aims at ensuring that individuals are able to get what they want and are willing to pay for, when they want it and are willing to pay for it – whether or not they act on integrated preferences.

I began work on this topic with the expectation that, in the absence of income effects, this willingness-to-pay measure of benefit would be equal to the area under the observed demand curve. My analysis has shown that, in general, that expectation is false. However, I believe that it has shown that the measure itself can be well-defined and capable of being calculated from observable data even when consumers do not act on integrated preferences. 'Specification 3' of the model of myopic choice presented in Section 4 is an illuminating test case. In this case, the area under the demand curve is 5, but the maximum yield of discriminatory pricing is 3. The difference between the two measures reflects the fact that, because the myopic consumer is never willing to pay for experiences that will occur in the future, there is no consumption profile for which the individual is willing to pay a total of more than 3. If the only real utility is that which people are willing to pay for, no consumption profile can be worth more than 3. I guess that some readers will conclude that this example illustrates the limitations of Dupuit's principle. But I am still inclined to endorse the principle and accept its implications.

References

Ariely, Dan, George Loewenstein, and Drazen Prelec (2003). Coherent arbitrariness: stable demand curves without stable preferences. *Quarterly Journal of Economics,118*, 73–105.

Auspitz, Rudolf and Richard Lieben (1889). *Untersuchungen über die Theorie des Preises* (Berlin: Duncker and Humblot).

Becker, Gordon M., Morris H. DeGroot and Jacob Marschak, (1964) "Measuring utility by a single-response sequential method", *Behavioural Science*, 9, 226-232.

Bernheim, Douglas and Antonio Rangel (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review: Papers and Proceedings* 97: 464–470.

Buchanan, James M. (1968). *The Demand and Supply of Public Goods*. Chicago: Rand McNally.

Buchanan, James M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Dupuit, Jules (1844/ 1952). On the measurement of the utility of public works. *International Economic Papers* 2 (1952): 83–110. Translated by R.H. Barback. First published in French in 1844. (Page references to reprint in *Cost-Benefit*

Ekelund, Robert and Robert Hébert (1985). Consumer surplus: the first hundred years. *History of Political Economy* 17: 419–454.

Hicks, John (1941). The rehabilitation of consumer's surplus. *Review of Economic Studies* 8: 108-116.

Houghton, R.W. (1958). A note on the early history of consumer's surplus. *Economica* 25: 49–57

Infante, Gerardo, Guilhem Lecouteux and Robert Sugden (2015). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, forthcoming.

Kahneman, Daniel and Amos Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263-291.

Kőszegi, Botond and Matthew Rabin (2007). Mistakes in choice-based welfare analysis. *American Economic Review* 97: 477–481.

Loomes, Graham, Peter G. Moffatt and Robert Sugden (2002). A microeconometric test of alternative theories of risky choice. *Journal of Risk and Uncertainty*, 24, 103–30.

Lowenstein, George (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology* 24 (4): S49–S56.

McQuillin, Ben and Robert Sugden (2012). How the market responds to dynamically inconsistent preferences. *Social Choice and Welfare* 38: 617–634.

Marshall, Alfred (1890). *Principles* of *Economics* (London: Macmillan).

Schelling, Thomas C. (1984). Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review* 74 (2): 1–11.

Slovic, Paul and Sarah Lichtenstein (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology* 78, 1–18.

Strotz, R.H. (1955-56). Myopia and inconsistency in dynamic utility maximisation." *Review of Economic Studies*, 23: 165-180.

Sunstein, Cass R. and Richard H. Thaler (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review*, 70: 1159–1202.

Sugden, Robert (2004). The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American Economic Review* 94: 1014–1033.

Sugden, Robert (2007). The value of opportunities over time when preferences are unstable. *Social Choice and Welfare* 29: 665–682.

Sugden, Robert (2013). The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed? *Inquiry* 56 (2013): 519–538.

Sugden, Robert and Alan Williams (1978). *The Principles of Practical Cost-benefit Analysis*. Oxford: Oxford University Press.

Tubaro, Paola (2006). Jules Dupuit's contribution to mathematical economics. Unpublished manuscript, Université Paris X. http://economix.fr/pdf/journees/hpe/2006-06-16_Tubaro.pdf.

Walras, Léon (1900/1954). *Elements of Pure Economics*, fourth edition. London: Routledge. Translated by William Jaffé. First published in French 1900. First edition 1874.