


ARTICLE

Voluntariness and the bounds of cost–benefit analysis

Robert Sugden 

School of Economics, University of East Anglia, Norwich, UK
Email: r.sugden@uea.ac.uk

(Received 31 January 2023; accepted 31 January 2023; first published online 27 February 2023)

Abstract

Behavioural economists often claim that their policy recommendations are justified by cost–benefit analysis (CBA), but without adequate explanation of the methodology they have in mind. I sketch the outlines of a CBA methodology that is compatible with the findings of behavioural economics and is in accord with my account in Sugden (2018) of a well-functioning market as a network of opportunities for mutually beneficial transactions. The key idea is that the CBA of a project is concerned only with effects that are not transmitted through voluntary interactions. I illustrate this proposal by considering the appraisal of fuel economy mandates.

Keywords: cost-benefit analysis; voluntariness; behavioural economics; regularisation; fuel economy mandates

Introduction

Behavioural economics needs a practical methodology for appraising public policies that does not assume that individuals' revealed preferences invariably satisfy the consistency axioms that are standardly used in neoclassical economics.¹ Two of the founding manifestos of behavioural welfare economics made the sweeping claim that policy advice could be based on some form of cost–benefit analysis (CBA), but were rather coy about what that form might be. According to Camerer *et al.* (2003):

A crucial assumption in our approach is that the bounds on rationality – their range and implications, as well as which policies help – are empirical questions subject to systematic analysis, and thus cost-benefit judgments can be made. (p. 1222)

Sunstein and Thaler (2003) made a similar claim:

How should sensible planners choose among possible systems, given that some choice is necessary? ... If feasible, a comparison of possible rules should be done

¹Neoclassical economics models consumers as rational maximisers of utility and firms as rational maximisers of profit. Behavioural economics relaxes these assumptions and uses some of the theoretical ideas and research methods of empirical psychology.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

using a form of cost-benefit analysis, one that pays serious attention to welfare effects. (p. 1190)

When I first read these claims, I was puzzled by the confidence with which they were made, and by the absence of supporting argument. On my understanding, the fundamental principle of CBA was the *potential Pareto improvement criterion* or *hypothetical compensation test*.² Or, more simply, CBA was about *willingness to pay*, as expressed by Dupuit, the mid-nineteenth century founder of CBA: ‘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’ (Dupuit, 1844/1952: 262). If, as behavioural evidence was showing was often the case, individuals’ preferences did not have the consistency properties that economics had traditionally assumed, how was the cost-benefit analyst to define and elicit preferences for use in CBA?

Since 2003, many behavioural economists have converged on a consensus about how that question should be answered. The essential idea is to decompose revealed preferences into two components – *true preferences* and *biases* (or *errors*). Then, at least if true preferences have the desired consistency properties, a behavioural CBA can be based on true preferences in the same way that traditional CBA was based on revealed preferences. With co-authors, I have argued that this approach is a blind alley: the concept of true preference lacks psychological foundations (Infante *et al.*, 2016). If that is right, the idea that an operational CBA methodology can be based on the distinction between true preference and error is a mirage. My aim in this paper is to take some preliminary steps towards developing a behavioural form of CBA in the tradition of Dupuit.

I must say straight away that I do not view CBA as an attempt to reach all-things-considered judgements about social welfare or the social good. I see it as a body of practical techniques that are used to produce a certain kind of summary information about policy options. In *Paretian* CBA (i.e. the traditional form of CBA that is based on the potential Pareto improvement criterion), the effect of a policy on each individual is defined as that individual’s *compensating variation*, i.e. the net change in money income that would exactly compensate them for the effect of the policy. ‘Compensation’ is defined using the neoclassical theory of preferences, i.e. as a money transfer that would keep the individual on the same indifference surface as if the policy was not implemented. The net sum of individuals’ compensating variations is the effect of the policy on *economic efficiency*. There is no claim that economic efficiency is the only relevant dimension that policy makers would or should consider. The background thought is more pragmatic – that economic efficiency is a meaningful concept, based on received economic theory, which provides useful information about the economic effects of a policy.

However, this conceptual framework rests on the assumption that individuals have consistent preferences over the possible effects of policies, so that compensating

²In a textbook about CBA that I co-authored in the 1970s, this principle was the defining characteristic of the ‘Paretian approach’ to CBA. My co-author was more inclined to allow a ‘social decision-maker’ to choose the objective of public policy, but saw the potential Pareto improvement criterion as a pragmatic first approximation to government objectives for the purposes of CBA (Sugden & Williams, 1978: 89–98).

variations are well defined. This assumption has been called into doubt by the findings of behavioural research. The issue I want to address is this: Is there a way of adapting the practices of Paretian CBA so that they do not require that assumption, but still produce useful summary information about the economic effects of policies?

The voluntariness boundary

I begin with a case study which at first sight may seem to have little to do with behavioural economics, but which introduces a key principle of the CBA methodology I am proposing.

In 1998, I was commissioned by the Department of the Environment, Transport and the Regions (DETR, the UK government department then responsible for transport policy) to recommend revisions to the official CBA methodology for appraising transport projects (Sugden, 1999). As viewed by DETR, the problem was to create an integrated appraisal system applicable both to road improvement projects and to public transport service improvements. These two types of project had very different economic characteristics. The use of the road system was essentially free of charge, and the construction of major roads was financed from general taxation; decisions about road-building were made directly by central government.³ Following the privatisations of the 1980s and 1990s, most public transport services were supplied by private firms, financed by a combination of fare revenue and government subsidies. Public decisions had to be made about whether to offer subsidies to firms for operating specific services. DETR wanted to create level playing fields between the appraisal of public and private transport projects, and between public transport operators (who were eligible for subsidy) and firms in other sectors of the economy (who in general were not).

I will focus on two common cases in which CBA includes an evaluation of effects on 'private' passenger transport users (i.e. individuals travelling to and from work or in their leisure time, as contrasted with in-work travel). The first case is a road improvement project. Typically, the predicted effect is a reduction in the time taken to travel on given links in the road network, and a consequent increase in the number of trips made on those links. The second case is an improvement in a public transport service, partly funded by an increase in government subsidy. Again, the predicted effects are changes in travel time and numbers of trips. Some of the effects of the service improvement are *user benefits* (or disbenefits), i.e. effects on the users of the service. But other effects are on the users of private cars. In urban settings, a major policy justification for subsidising public transport is to induce a switch from cars to public transport, thus reducing overall traffic congestion.

At the time, DETR and the passenger rail regulator (the Office of Passenger Rail Franchising) took the view that (with some exceptions) user benefits should *not* count as a justifications for public transport subsidies because they were not externalities to the activity being subsidised. In contrast, indirect benefits to private car users

³Fuel used in road transport was subject to a high level of duty, and all motor vehicles used on public roads were subject to a lump-sum 'road fund tax'. How far those taxes should be interpreted as charges for road use, analogous with public transport fares, was (and remains) unclear.

should count. In my report, I presented an economic justification for this type of *restricted CBA* (i.e. CBA which is restricted to the measurement of external effects), based on two main arguments.

The first argument was about maintaining a level playing field between public transport operators and other firms. Public transport improvements create user benefits through increases in consumers' surplus, which arise because most consumers of public transport would be willing to pay more than the prices of the services they consume. If perfect price discrimination were possible, an operator would be able to capture the added value to consumers of a service improvement. Of course, perfect price discrimination is an impossible polar case, but that is true in all sectors of the economy. It is not a normal practice to use public funds to compensate firms for failing to fully capture consumers' surplus.

The second argument went deeper. A fundamental aspect of the motive power of a market economy is that firms are rewarded for discovering mutually beneficial transactions with consumers. Thus, it should be each firm's responsibility to find ways of inducing potential customers to cover the cost of producing goods that those customers want to buy. Price discrimination is one of those ways.⁴

To understand the logic of restricted CBA, it is necessary to recognise a distinction, fundamental to CBA, between *pecuniary* and *technological* externalities. Generically defined, an *external effect* of an activity is a cost or benefit of that activity that is experienced by someone who has not chosen to participate in it. Pecuniary externalities are a subcategory of external effects. These are gains and losses that are experienced by non-participants as a result of price changes induced by the activity and transmitted by the workings of the market. For example, consider an improvement to an urban rail service whose direct effect is to reduce travel times for users. Suppose this effect induces an increase in the rental price of housing close to train stations. This price increase is a negative pecuniary externality to buyers in the rental market and an equal and opposite positive pecuniary externality to sellers. Since the two effects cancel out when compensating variations are summed, they do not affect the result of a Paretian CBA. In combination, they constitute a redistribution of some of the benefits of the rail improvement from rail users to property owners. In contrast, if the rail improvement reduces travel time for road users, this is a 'technological' external effect – an effect that is *not* transmitted through induced changes in market prices. A Paretian CBA should include such effects.

In this sense, Paretian CBA is *bounded by the market*: in appraising a project, there is no need to take account of effects of that project that are transmitted through markets. A CBA methodology that recognises this boundary can be construed as respecting two default presumptions of economic policy in a social market economy – that governments should not second-guess the private decisions of firms and consumers, and that issues of income redistribution should be separated from those of microeconomic management.

I have argued elsewhere that a behavioural defence of a social market economy can start from a conception of a competitive market as a domain of voluntary

⁴This argument can be traced back to Dupuit (1844/1952). It is developed in more detail in Sugden (2018: 160–164).

interactions, without any reference to individuals' supposedly 'consistent' or 'true' preferences (Sugden, 2018). In the present paper, I explore the idea that behavioural CBA should be bounded by the *domain of voluntary interaction* – that it should take no account of effects that are transmitted through voluntary interactions.

Applying this idea to the case of a subsidy to a public transport operator, the taxpayers' role in funding the subsidy is not voluntary, and so the subsidy should count as a cost in the CBA. In contrast, the relationship between the government and the operator is voluntary. The government offers to pay the operator to provide a particular service improvement, and the operator accepts the offer. Similarly, the operator voluntarily chooses to change the offers it makes to customers, and potential customers voluntarily choose whether to accept or reject these offers. On the account that I am proposing, it is not the role of CBA to ask whether the decisions of the operator and the customers are rational or reasonable. However, the relationship between users of public transport and users of private cars is an externality, not a voluntary interaction: the car users benefit from reduced road congestion without being required to pay for it. So this effect counts as a benefit in the CBA.

Regularisation

In the case of the subsidised public transport service, the government is acting on behalf of the car users who stand to gain from reduced road congestion. It is bringing about the supply of a good (less congested roads) which citizens value but which is not bought and sold on the market. The cost of supplying this good (the subsidy) is being charged to citizens in the form of tax payments. The purpose of CBA is to assess whether the benefit of less congested roads is greater than the cost of the subsidy.

A crucial step in the CBA is to convert savings in travel time to monetary values. At the time of my report, the UK government used a single *equity value* of time to evaluate the benefits of road improvement projects. This was a weighted average of *behavioural values* of time derived from econometric studies of individuals' choices between travel modes. These studies estimated models in which rational individuals made trade-offs between travel time and money cost. The behavioural value of time for a given class of individuals (for example, classified by income) was the value that best explained the totality of the decisions of those individuals.

One issue that I considered in my report is illustrated by the following example. Suppose a private firm proposes to operate a new public transport service that caters to a niche market of consumers who are willing to pay unusually high prices for speed and convenience. To be more concrete, suppose the proposal is for a high-speed train service between an international airport and a city centre. The potential customers' high willingness to pay for time savings might be explained by the high incomes of international air travellers. Or it might be explained by the psychology of context-dependence. Instead of considering the costs and benefits of the train trip in isolation, these customers may think of it as a small component of an international journey; in that context, the rail fare is perceived as a minor add-on. Whatever the reason, the firm's market research supports the expectation that the service will be profitable.

Suppose the project cannot go ahead without government approval, and is appraised by a cost–benefit study. This study finds that the project will have some technological external effects in the form of changes in travel time on different links of the road network, but these costs and benefits are relatively minor and, when valued at the equity value of time, sum to approximately zero. Should the project pass the CBA test?

If user benefits were measured using the equity value of time, the project might fail the test. But would that be a reason for the government to refuse approval for the project? My answer was ‘No’. According to the criterion I presented in the previous section, the interaction between the firm and its customers is located in the domain of voluntary interaction, and therefore outside the scope of CBA. In principle, the non-voluntary effects *are* relevant, but their net cost is close to zero.

Nevertheless, it might be objected that the equity value of time was derived by estimating the parameters of a model of rational choice. In the CBA of road improvements, it was treated as the value of every citizen’s travel time. So why should it not also be used in measuring the user benefits of a profitable public transport project?

The answer is that the equity value of time was not the *truth* about citizens’ preferences between time and money; it was a *regularisation* of disparate observations of actual behaviour.⁵ The model that was fitted to those observations was just a convenient general-purpose model that happened to use concepts taken from a theory of rational choice. Because reductions in road congestion are not the objects of voluntary transactions between individuals, the appraisal of road improvements requires time savings to be valued, and if those valuations are to be responsive to individuals’ preferences, some form of regularisation is unavoidable. But there is no need to second-guess individuals’ actual decisions about whether to participate in voluntary interactions.

An example: fuel economy mandates

As a further illustration of my approach, I consider a case that Sunstein (2014, 2021) has used as a leading example of behavioural CBA – the appraisal of fuel economy mandates, as set by the US Environmental Protection Agency (EPA).

Corporate average fuel economy standards are legal restrictions imposed on US car manufacturers, imposing constraints on the proportions of different models produced. The aim is to increase the proportion of fuel-efficient cars in the product mix. The primary aim of these regulations is to reduce exhaust emissions, for reasons of public health and the mitigation of climate change. It is uncontroversial that a CBA of this type of regulation should count reduced emissions as a benefit: the government is bringing about an increase in a public good (low emissions) which citizens value. But, over and above the value of reduced emissions, should drivers’ reduced expenditure on fuel count as a benefit? This is an important question. In a CBA of fuel economy standards conducted by the US Department of Transportation, fuel cost savings *were* included, and accounted for 84% of total benefits (Sunstein, 2021: 498–499).

⁵The distinction between ‘regularising’ and ‘purifying’ (or ‘debiasing’) revealed preferences was introduced by Infante *et al.* (2016). See also Sugden (2022).

According to Sunstein:

[C]onsumer savings from fuel-efficient cars and from energy-efficient appliances unquestionably count as benefits. They represent savings to consumers, brought about by regulation. The hard question is not whether they count as benefits, but the identification of the relevant market failure. (2014: 141)

It seems to me that the first two sentences make economic sense only if one assumes that the two types of car differ *only* in terms of fuel consumption. (Compare a regulation which prohibits the sale of high-quality, high-price varieties of cheese. Consumers spend less on cheese. Is that unquestionably a benefit of regulation?) Identifying the ‘relevant market failure’, Sunstein endorses the view of the EPA that fuel economy mandates counter an array of behavioural market failures, i.e. patterns of consumer behaviour which reveal error or bias relative to a standard of ‘true’ preference or welfare.⁶ These alleged failures include myopia, loss aversion, lack of information, lack of understanding of information and insufficient attention to relevant information; they combine to create the ‘energy paradox’ that ‘consumers appear not to purchase [energy-efficient] products that are in their economic self-interest’ (Sunstein, 2014: vi).

Sunstein recognises that some of the patterns of behaviour that he classes as market failures are ones that the consumers themselves, even on careful reflection, would not acknowledge as mistakes or would want to correct. For example, he notes that many consumers are reluctant to pay premium prices to buy cars which embody unfamiliar technologies ‘even when such vehicles have attractive EPA fuel economy ratings’ (Sunstein, 2021: 503). As viewed by the consumer, this attitude to new technologies is surely caution rather than error. As evidence of the energy paradox, Sunstein (2021: 502) cites an econometric study that investigated how car purchasing decisions in the US were affected by a ‘significant correction of an erroneously stated miles per gallon measure’. On the assumption that car buyers fully believed both the pre-correction and post-correction measures, the researchers estimated implicit trade-offs between car prices and the discounted cost of fuel consumption. The headline finding was that car buyers were willing to pay a price premium of between \$0.15 and \$0.38 for fuel savings which, according to the official ratings, were worth \$1 (Gillingham *et al.*, 2021). There is unintended irony here: the correction was of previously EPA-approved fuel economy ratings which, after an EPA audit, had been found to be based on significant (but supposedly unintentional) over-statements by the manufacturers. Perhaps part of the explanation of the energy paradox is that consumers are understandably sceptical about the trustworthiness of fuel economy ratings.

To show how my kind of CBA would appraise fuel economy mandates, I present a very simple model of the working of such a mandate and a CBA that applies to that model.

⁶Sunstein is ambivalent about whether the standard is of preference or welfare. He consistently maintains that the standard should be ‘as judged by [the individuals] themselves’, but sometimes the component of individual judgement is very thin – as in his proposal ‘to ask about what would increase people’s aggregate welfare over time, on the theory that aggregate welfare (taking all relevant values into account) is the end that people really do care about’ (Sunstein, 2014: 70).

Suppose there are just two types of car. ‘Type 1’ is *fuel-inefficient*; ‘Type 2’ is *fuel-efficient*. In the *do-nothing* scenario, the two types are marketed at prices p_1 and p_2 , with $p_2 > p_1$. (Think of these as per-period prices of rental contracts between consumers and manufacturers.) In the *mandate* scenario, the government fixes a minimum proportion of type 2 cars in the product mix (greater than in the *do-nothing* scenario) and requires each firm’s production levels to meet that requirement. Prices must adjust so that consumers choose to buy the mandated proportions of cars. In the mandate scenario, the prices are $p_1' > p_1$ and $p_2' < p_2$.

Emissions are lower in the mandate scenario, and this clearly counts as a benefit. (I bracket out the question of how it should be valued.) This benefit has been achieved through a governmental intervention in the car market which was not part of a voluntary transaction, and therefore outside the voluntariness boundary. A CBA needs to take account of the effects of this intervention on buyers and sellers in that market. For car manufacturers, the effect is a change in profit. For consumers, the effect is the change in the two prices. There is a loss of consumers’ surplus in the type 1 market (in which the price increases from p_1 to p_1') and a gain in the type 2 market (in which it falls from p_2 to p_2'). If the prices and quantities traded in the two scenarios are known (or predicted), measuring changes in consumers’ surplus is straightforward.⁷

What about changes in expenditure on fuel? Contrary to Sunstein’s argument, reductions in expenditure on fuel do *not* count as benefits. The *involuntary* effect of the mandate is on the market for cars. The changes in car prices induce changes in consumers’ demand for fuel, but this effect is transmitted through voluntary transactions in the market. Viewed in the conceptual framework of CBA, as that has been understood since Dupuit, this is not an additional cost or benefit of the price changes; it is one of the ways in which the market adjusts to those changes.

My CBA accounting draws attention to some significant distributional effects of fuel economy mandates. The mandate imposes losses on consumers who would buy fuel-inefficient cars irrespective of whether the mandate was imposed: they have to pay more to buy goods that they want. It confers corresponding benefits on consumers who would buy fuel-efficient cars irrespective of whether the mandate was imposed. If the purpose of the mandate is to reduce emissions, these distributional effects can be seen as fair results of the ‘polluter pays’ principle. Sunstein (2014: 115) defends fuel economy mandates as showing respect to freedom of choice by giving ‘a great deal of freedom and flexibility to both companies and consumers’. The idea, I take it, is that the government is requiring the participants in the car market to bring about a specified reduction in emissions, but is allowing them to choose how to do this. Consumers who do not contribute directly to the reduction by using fuel-efficient cars contribute indirectly, by paying more for their cars, while those who contribute directly are partially compensated by paying less for theirs. But this logic breaks down if the mandate is interpreted as a policy to counter behavioural market failures.

⁷The reader may ask whether, in using the concept of consumers’ surplus, I am smuggling in questionable neoclassical assumptions. In fact, the assumptions I need are quite weak. Consider a consumer whose opportunity set O is defined by some vector \mathbf{p} of market prices and an income constraint; the consumer chooses some quantity vector \mathbf{q} from O . Suppose there are very small changes in some of these prices. The key idea behind consumers’ surplus is that the consumer’s net gain from these price changes can be approximated by the net reduction in the cost of buying \mathbf{q} .

To the extent that the mandate is interpreted in this way, the loss imposed on consumers who continue to buy fuel-inefficient cars is not a case of a polluter paying for harm inflicted on others; it is an additional harm imposed on people who are already supposedly harming themselves. Conversely, the gain enjoyed by consumers who would anyway buy fuel-efficient cars is not a return for any benefit they are conferring on others. Such arbitrary gains and losses are not signs of freedom and flexibility. They might be better described as invitations to rent-seeking.

Conclusion

In this paper, I have tried to sketch the rough outline of a methodology of CBA that is compatible with the findings of behavioural economics and in accord with my account, presented in Sugden (2018), of a well-functioning market as a network of opportunities for mutually beneficial transactions. I commend this form of CBA, not as telling public decision makers what they ought to do, but as a way in which economists can contribute their professional expertise to a democratic process of public decision making.

References

- Camerer, C. F., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003), 'Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism"', *University of Pennsylvania Law Review*, **151**: 1211–1254.
- Dupuit, J. (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 Analysis*, edited by Richard Layard. Penguin, 1973.)
- Gillingham, K., S. Houde and A. Benthem (2021), 'Consumer myopia in vehicle purchases: evidence from a natural experiment', *American Economic Journal: Economic Policy*, **13**(3): 207–238.
- Infante, G., G. Lecouteux and R. Sugden (2016), 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', *Journal of Economic Methodology*, **23**: 1–25.
- Sugden, R. (1999), 'Developing a Consistent Cost-Benefit Framework for Multi-Modal Transport Appraisal'. Report to Department of the Environment, Transport and the Regions. Retrieved from: <https://research-portal.uea.ac.uk/en/publications/developing-a-consistent-cost-benefit-framework-for-multi-modal-tr>
- Sugden, R. (2018), *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.
- Sugden, R. (2022), 'Debiasing or regularisation? Two interpretations of the concept of "true preference" in behavioural economics', *Theory and Decision*, **92**: 765–784.
- Sugden, R. and A. Williams (1978), *The Principles of Practical Cost-Benefit Analysis*. Oxford: Oxford University Press.
- Sunstein, C. (2014), *Why Nudge?* New Haven, CT: Yale University Press.
- Sunstein, C. (2021), 'Behaviorally informed mandates? Internalities, externalities, and fuel economy rules', *N.Y.U. Environmental Law Journal*, **29**: 493–505.
- Sunstein, C. and R. Thaler (2003), 'Libertarian paternalism is not an oxymoron', *University of Chicago Law Review*, **70**: 1159–1202.

Cite this article: Sugden R (2023). Voluntariness and the bounds of cost-benefit analysis. *Behavioural Public Policy* 7, 954–962. <https://doi.org/10.1017/bpp.2023.7>