Timothy J. Brennan*

Behavioral economics and policy evaluation

Abstract: Behavioral economics posits a number of cognitive biases and limitations, which raises questions as to whether revealed willingness to pay equals true willingness to pay. If so, benefit-cost analysis, with a number of methodological advantages, would need to be replaced. Prior analyses of the issue by Sunstein, Sugden, and Bernheim and Rangel fail to offer guidance that would avoid substituting centralized judgments for decentralized information on benefits and costs. Alternatives including using post-implementation valuations, libertarian paternalism, and direct democracy on policy issues also have conceptual or practical limitations. A tentative suggestion is democratic delegation, somewhat appealing because it is already applied to cope with bounded rationality and non-efficiency values. Viewing benefit-cost analysis as a market analogue, and restricting the domain of behavioral economics to uninformed consumers, may be useful guides. The most important guidance may be to require very strong evidence of substantial choice failure before abandoning benefit-cost analysis.

Keywords: behavioral economics; benefit-cost analysis; collective decision-making; program evaluation.

DOI 10.1515/jbca-2013-0022

1 Introduction

Behavioral economics (BE) – which we might define as the analysis of decisions contrary to the assumption that people make choices that best promote their self interest as they see it – is having a growing role in policy assessment.[1] Exam-

1 Richard Thaler, a leading developer and proponent of behavioral economics, calls it "the combination of psychology and economics that investigates what happens in markets in which some of the agents display human limitations and complications" (Mullainathan & Thaler, 2001, p. 1094).

***Corresponding author: Timothy J. Brennan,** University of Maryland, Baltimore County – Public Policy 417, Public Policy Building UMBC, 1000 Hilltop Circle, Baltimore, MD 21250, USA; and Resources for the Future, 1616 P St. N.W., Washington, DC 20036, USA, Tel.: +1-410-455-3229, Fax: +1-410-455-1172, Tel.: +1-202-328-5084, Fax: +1-202-939-3460, e-mail: brennan@umbc.edu

ples abound. Much energy policy rests on a premise that consumers and business owners fail to invest in more efficient lighting, heating, and cooling, when the value of energy saved exceeds the cost of those investments at any plausible discount rate. A manifestation of this in climate policy is the expectation that a considerable amount of greenhouse gas mitigation can be achieved at "negative cost," reflecting the idea that the benefits of programs exceed costs even if they have not yet been adopted, without taking carbon mitigation benefits into account (Enkvist, Nauclér, & Rosander, 2007). Another prominent policy example is that whether people enroll in employer-subsidized defined contribution pension plans depends on the apparently irrelevant and trivial fact of whether they have to check a box to opt into the plan, or check a box only if they want to opt out of a plan (Choi, Laibson, Madrian, & Metrick, 2002).

Moving past standard economics to behavioral economics to explain behavior in laboratories or practical settings raises important methodological questions. Our focus here is policy evaluation. Specifically, behavioral economics challenges the foundations of benefit-cost analysis (BCA). A defining presumption of BCA is that one can estimate how much people are willing to pay for the benefits of a policy, or to avoid the costs of a policy, from observed demand and supply. While making those observations presents major and familiar empirical challenges, the premise of the exercise is clear – that we can infer value of something to people from their revealed preference for it. Behavioral economics rejects that premise, and thus compels us to find some other means of estimating benefits and costs.

Any alternative will require that one has a way to ascertain or impute "real" preferences, if one cannot infer them from choices. The challenge will be to see if there is a way to come up with real preferences that avoids paternalism or substitution of the policy makers' preferences for those of the individuals affected by the policy choices. If behavioral economics becomes an important consideration in explaining those choices, this challenge will have to be met.

We begin by looking at five methodological justifications for BCA, at least in its ideal form. To the extent that BCA constitutes acceptable practice in policy evaluation – and we recognize that it is not always accepted as implemented – it provides a useful standard against which evaluations incorporating behavioral economics considerations can be judged. In particular, one ought not hold BE accountable for failing to meet standards that idealized BCA does not meet. We then see how BE threatens these advantages.

This leads us to ask how one could design an evaluation method with the properties of BCA but consistent with the methodological premises of BE in contrast to standard economics. A first step in that regard is to examine Sunstein's (2000) argument that BE justifies the employment of BCA. It turns out that his

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

discussion is about justifying the substitution of government expert judgment for democratic decision-making because of the latter being prone to some of the decision-making errors that BE has identified. Without failures of democracy, it is difficult to justify the use of BCA in making policy choices rather than deferring to the outcome of the policy process even without BE's challenges to the neoclassical paradigm. However, Sunstein's assessment leaves open the question here as to how BCA should be conducted when the presumptions underlying its typical measures of benefits and costs are, according to BE, no longer valid. We also examine for possible guidance Sugden's (2008) defense of markets despite cognitive biases BE identifies and his (2005) proposal to base benefit-cost analysis on the willingness to pay that a person would express were a market to simulate the policy choice at hand. We then turn to Bernheim and Rangel's (2009) effort to derive formally economic welfare measures consistent with BE.

We take a look at a variety of options. One would be to use central policy judgment instead of market data alone to determine true costs and benefits, calculating benefits and costs with those data rather than data derived from the market. This requires resolution of the question of how one knows that right answer. Thaler and Sunstein (2003, 2009) propose "libertarian paternalism," in which policy makers only change default choices toward what they believe to be the best options, but where persons retain the ability to overrule the expert judgment at little cost. As noted above, a leading example is requiring employers to offer "opt-out" rather then "opt-in" pension contribution plans. This option applies only in limited contexts, and suffers from the conceptual error in assuming that the default choice conveys no information that explains (if not justifies) the propensity of people to choose that default.

A third option would be to apply benefit-cost analysis ex post rather than ex ante, seeing if the winners could compensate the losers after the policy has been invoked and, presumably, people are no longer seeing the errors that led would have led to a failure to adopt the policy (Brennan, 2010). The difficulty here is error – what happens if the policy fails and is difficult to reverse? This also forces us to consider, fourth, how one would evaluate policies if the policies involve preference change. A fifth policy would be using democratic institutions rather than technocratic standards to make the choices, but as Sunstein suggests, those specific choices are likely to be problematic as well if one is taking BE seriously in policy evaluation.

This leaves democratic delegation, in which the voters select not policies but policy makers on the basis of when they would reject willingness to pay (WTP) data in choosing policies and how they would make those choices in light of that rejection. Such a policy could justify a host of policies already in place, e.g., having the FDA decide when a drug should be available only with prescription

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

to save consumers the effort of and potential error in evaluating risks and benefits on their own. Such delegation also replicates the five rationales for adopting idealized benefit-cost analysis as a policy standard. However, such choices may also be subject to behavioral error. More importantly, democratic delegation continues to beg the question as to what these competing policy makers should offer the public in seeking the authority to make decisions on the public's behalf. A number of suggested methodological principles to guide these decision makers are offered. Whether this generates an acceptable policy evaluation mechanism consistent with behavioral economics or responsive to its challenges remains an open question.

2 Methodological advantages of BCA

To identify and critique policy evaluation tools that may be consistent with BE, we should start with looking at the desirable properties for the policy evaluation tool consistent with conventional economics – benefit-cost analysis. It would be unreasonable to downgrade BE-based methods for failing to meet standards that BCA does not meet. To the extent that BE-based methods exceed BCA by those criteria, or satisfy additional criteria, their advantages would be clarified as well.

As just noted, in making these comparisons, we focus primarily on the idealized application of BCA. Its shortcomings in practice have been widely noted, with three most compelling.[2] First, shortcomings in the data, and the attendant range in judgment in interpreting data in terms of benefits and costs, create sufficient scope for differences in outcomes that BCA can be amenable to tailoring by policy advocates rather than serving as an objective means for resolving disputes among those advocates. Second, the cost of doing BCA, and that the burden of proof generally rests on those who want to impose a regulation, may combine to bias the policy process against the imposition of regulations. Third, the theoretical claim that BCA is agnostic regarding distributional outcomes is belied by the general absence of mechanisms through which winners compensate losers.

Whether these flaws are substantively valid is a question outside the scope of this paper. For now, we will generally look at BCA as if it were practiced without data shortcomings, advocacy manipulation, institutional inertia, or inattention to compensation.

1. *Connection of measures to benefits and costs*. Any other virtues of BCA matter little if at all if the inputs into the calculation have little to do with

2 Revesz and Livermore (2008) reviewed such criticisms in a defense of BCA.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

what we might call benefits and costs. To the extent the costs and benefits are uncontroversially monetary – either determining profits reaped by presumably profit-maximizing firms or reaped or borne by individuals in conventionally monetary contexts, such as wages, prices, or investment income – the monetary measures used in BCA match these benefits pretty well. The more difficult justification of the correspondence comes at the terminals of the chain from production and consumption, which begin and end with how much people have to be paid to give up their time, labor, and property, and how much they are willing to pay to acquire goods and obtain services. Standard economics takes this as a relevant if not unique measure of benefits and costs, partly for reasons discussed below. Also discussed below is that the root challenge of BE to BCA involves the validity of this assumption.

2. *Decentralized measure of benefits and costs.* A first advantage of BCA is that it is based on decentralized measures of benefit. The evaluation data are "out there," in principle exogenous to the policy decision process. Ideally, all of those interested in a policy outcome can obtain the same evidence on its costs and benefits, derived ultimately from demand curves on the benefit side and cost or supply curves on the cost side. Whether supply curves can be estimated and convey cost information depends on whether the underlying markets are competitive, certainly a practical complication.
3. *Interpersonal comparability.* A useful measure for assessing policy has to be able to take into account the inevitable disparities in the magnitude and direction of the effect of policies on everyone affected. Some win a lot, some win a little; some lose a little, some lose a lot. Reporting the effect of a policy as a vector with N coordinates, reflecting the net effect on each of N individuals, is going to be of little to no use in comparing two policies, or even whether to enact a policy or to do nothing. BCA overcomes this problem crudely but directly: Add up the net effects revealed by demand and cost curves across all individuals; if the number is positive, the benefits exceed the costs. Apart from whether this is normatively justified, the process is (with ideal data) feasible and, as it utilizes decentralized data, produces a result independent of the analyst's policy preferences.
4. *Pareto feasibility.* A leading complaint about BCA is the utilitarian one that a dollar of benefit or cost to one person may carry considerably more ethical weight than a dollar of benefit or cost to another person. The usual example is that a dollar to a very poor household is ethically more valuable than a dollar to a billionaire. The standard response, that allows BCA users to transcend this problem, is to observe that if the benefits exceed the costs, the winners in principle can compensate the losers, leaving everyone better off. Whether

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

policy makers decide to make everyone better off or accept the distributive consequences of the policy is not the problem of the benefit-cost analyst.

There are at least three notable shortcomings to this argument. One, more theoretical in standard economic terms than frequently empirical, is that the valuations can be sensitive to whether or not the policy is enacted, creating the possibility that winners could compensate losers in choosing *A* over *B* and in choosing *B* over *A*, depending on the starting point, leaving no unambiguous policy recommendation.[3] Second is that the theoretical compensation by winners of losers may not be feasible in practice; for example, it may be impossible to identify each winner and loser, ascertain benefits and costs, and take from the former group to give to the latter. Third, partly because of this prior difficulty and, more likely, political realities, the losers in a choice of policy *A* over *B* may predictably fail to be compensated by *A*'s winners. To ignore that reality on the grounds that winners could in principle compensate losers may be ignoring morally relevant advice. But, again ideally, the possibility of "win-win" under BCA is an important desirable property. To put it conversely, an evaluation method that inherently fails to do so, or fails to aspire to do so, seems more problematic at the outset.

5. *Internal consistency.* A final appealing aspect of BCA worth noting separately is the internal consistency across all four of the prior properties. It measures something at least related to benefit and costs, and that thing it measures, added across everyone yet based on decentralized, individual revealed preferences, is also the mechanism by which compensation in principle could be implemented to make everyone better off. This is a special case of a larger property of conventional economics: The means for ascertaining preferences, measuring incentives and describing behavior on the positive side also supply the efficiency norm by which institutions may be compared with regard to their ability to give people more preferred bundles of goods and services.[4]

3 As noted below, BE findings suggest that such reversals may be more frequent and profound.

4 It is important to be clear that the consistency between the two does not make the positive analysis normative. One could point out that one outcome is more efficient than another, or even Pareto-preferred without claiming that the more efficient outcome be realized. But simply as an observation from teaching in a multidisciplinary department, it has long seemed to me that a methodological distinction of economics compared to other social sciences in analyzing policy is this consistency between the positive and normative side, at least as compared with political science and sociology, however valuable are the positive insights and normative cautions contributed by these other disciplines.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

3 BE as threat to all five

The challenge from BE comes about essentially because it raises questions about whether any of these five desirable procedural and substantive properties of standard BCA continue to hold.

1. *Connection of measures to benefits and costs.* The BE challenge to BCA begins here. The fundamental claim of behavioral economics is that people's choices can be mistaken, in the sense that their choices do not reflect what they really want. Because of a host of biases and cognitive limitations,[5] people's decisions need not represent their "true" preferences. The implication for BCA is that the willingness to pay revealed in choices and measured by the areas under demand curves or over supply curves do not reflect the actual willingness to pay. Consequently, the measures underlying BCA do not correspond to actual benefits and costs.

   The problem is more fundamental than inaccuracy – it is that different decision contexts can give different answers. Bernheim and Rangel discuss this in abstract terms, but it may be helpful to look at a specific example – the endowment effect combined with loss aversion (Knetsch 1989; Kahneman, Knetsch, & Thaler, 1991).[6] A familiar manifestation of this phenomenon in BCA is the large divergence between what people say they are willing to pay to achieve a particular outcome and what people say they would have to be paid to accept the loss of that outcome (Horowitz & McConnell, 2002). The central idea is that my willingness to pay to have *A* instead of *B* depends on whether I start off with *A* or with *B*. One would observe some difference, for example, a higher willingness to pay for *A* starting with *A* than with *B* if

5 We do not normally recommend citing Wikipedia, but the entry "List of biases in judgment and decision making" contains over 100 different supposed biases in decision making and belief formation. http://en.wikipedia.org/wiki/List_of_cognitive_biases.

6 The term "loss aversion" here requires some care. It often is presented in shorthand as saying that an individual is more averse to a loss of $X than a gain of $X. However, that is the definition of risk aversion in standard economics. "Loss aversion" in behavioral economics requires an endowment effect. Using the terminology in the text, it says that if someone prefers *A* to *B*, her willingness to pay for the "gain" of *A* starting at *B* is less than her willingness to pay to avoid the loss of going back to *B* if she starts at *A*, or to put it in BE terms, views *A* as part of her endowment. One might view the goal of advertising as not to persuade a consumer that *A* is better than *B*, but to convince a consumer that *A* is part of her endowment, what she should regard herself as having. If so, she will be willing to pay more to keep it by buying it than she would to add it if she regarded her endowment as including only *B*. The added and more controversial claim with loss aversion on top of the endowment effect is that people are risk preferring when it comes to avoiding losses, even if risk averse when it comes to evaluating gains.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

having *A* conveys additional wealth and *A* is a normal good, that is, one for which the willingness to pay increases with income.

The possibility of income effects inducing different outcomes in BCA – the Scitovsky paradox – is well known. However, the differences in many cases appear to be much greater and more frequent than what a mere income effect would predict. If one gets different measures based upon starting points or external effects that should not matter to underlying benefits and costs, the measures themselves are not just in error in some stochastic sense. They become questionable if not intrinsically useless as measures in principle.

2. *Decentralized measure of benefits and costs.* If BE is correct that the most accurate measures of revealed willingness to pay fail to reflect actual benefits and costs, then determining those actual benefits and costs will require adjustments for error. Adjustments require an adjuster and a procedure for adjustment. In some settings, discussed below, one might be able to make a persuasive case that the error might be based on decentralized data. For example, one might argue that that data indicate that an individual's choice is anomalous, such as reflecting a discount rate out of line with most of her other decisions. One could then correct the anomalous decision by using the discount rate consistent with the other decisions. This still requires that the outside analyst have (and impose) a reason for believing that the anomaly is a real one rather than a rational decision reflecting particular preferences reflecting that specific setting. The implication is that with BE, the analyst must actively determine which choices accurately reflect preferences, which do not, and how the latter should be adjusted to measure actual benefits and costs.[7]
3. *Interpersonal comparability.* A controversial hallmark of BCA is its treating dollars of gains and losses the same, regardless of who reaps or bears them. The normative acceptability of such controversy comes from being able to reconcile gains and losses. But to be able to make a comparison, one has to have sufficiently stable measures to begin to justify comparisons. One need not have disaggregated data; the policy analyst can exercise her judgment in making interpersonal comparisons of benefits and costs. But the task seems not easy if the underlying data either reflect consumer error or depend on contexts that would seem to have limited if any relevance to the interpersonal comparisons.

7 Of course, conventional BCA requires a claim that there are no biases distorting the benefit and cost information contained in empirically observed demand and cost curves. To be clear, the claim here is not that BE is incorrect, but that if it is correct, the advantages of BCA fail to hold, requiring investigation of replacement methods for policy evaluation.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

4. *Pareto feasibility.* The normative usefulness of BCA springs from the idea that because of comparability, an excess of gross benefits over gross costs implies that in principle the winners can compensate the losers. As noted above, even under standard economics, this principle may not hold when income effects are consequential. Young (1993, pp. 12–14) has offered the example of the allocation of a single kidney to one of two who need it; presumably the winner will be unable to compensate the loser if the loser will die as a result. However, if endowment effects make this more widespread, the possibility of compensation will shrink, eliminating the ability of the policy analyst to regard distributional considerations as outside the scope of his method.
5. *Internal consistency.* With no clear measure attached to benefit because of error or contextual effects, and no ability to make interpersonal comparisons, the ability to have a consistent method for both measurement and compensation evaporates. One might be able to restore consistency by dropping the other properties. The tempting property to drop will be decentralization: The policy analyst can make her judgments on benefits and costs, defining a Pareto improvement as she views appropriate. Whether one is left with centralized judgments rather than decisions based on observed consumer behavior is the primary challenge BE presents for policy assessment. Before looking at how that might be assessed and implemented, we turn to three leading views that speak to whether and how one might define policy standards in a world of consumer error and contextual effects.

4 Three potential approaches

Among those who have addressed the policy relevance of behavioral economics, particularly with respect to BCA, three merit mention: Sunstein (2000), Sugden (2008), and Bernheim and Rangel (2009).

4.1 Cass Sunstein (2000)

Sunstein's experience on this topic is considerable, not just as an academic student of both behavioral economics and benefit-cost analysis, but as a practitioner. From 2009 to 2012, he served as the director of the Office of Information and Regulatory Affairs (OIRA) in the Office of Management and Budget, essen-

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

tially the person in the Executive Office of the President empowered to review Federal regulations. Since President Obama has largely continued Executive Orders issued by Presidents Reagan and Clinton supporting a requirement that benefits at least justify costs,[8] Sunstein's job entailed the review of benefit-cost analyses undertaken by Federal agencies.

Sunstein has argued that BE and BCA are complements rather than substitutes, that the cognitive failures identified by BE strengthen the case for BCA. As this largely appears to contradict most of what we have argued so far, and because of Sunstein's practical experience as well as academic expertise, his argument presents an important challenge warranting careful examination.

The central issue in Sunstein's argument is not about what BE says about how to do benefit-cost analysis, but whether to do "it"; the scare quotes to be explained. The reason to do BCA or something like it is because some of the cognitive limitations and biases BE has identified imply that democratic processes will lead to mistaken electoral choices. The specific biases he identifies are:

- *Availability heuristic*: People overestimate risks from publicized events (hurricanes, plane crashes) and downplay risks that receive less attention.
- *Dangers "on screen"*: People may overestimate risks that are in front of them and downplay countervailing benefits that may not be as obvious.
- *Systemic effects*: People may make decision based on near term effects and neglect systemic effects that may occur later on.
- *Alarmist bias*: Newly discovered risks may get too much weight, even if the discovery includes information that the risk is low.
- *Separate evaluation*: A person's willingness to pay to avoid a risk may be considerably greater if viewed in isolation (e.g., preventing breast cancer) than as part of a larger whole (of one's willingness to pay to prevent all cancer, how much of that amount applies to breast cancer, colon cancer, prostate cancer, etc.).

Sunstein argues that errors arising from these biases can be "aggravated" by information cascades. An information cascade arises when random or erroneous choices made by a few people early on lead others to disregard their own information, assuming that the early deciders knew what they were doing. They may also not want to suffer the reputational consequence of going against an apparent consensus. Information cascades themselves are consistent with how individuals

8 Executive Office of the President, Executive Order 13563, "Improving Regulation and Regulatory Review," issued January 18, 2011.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

process information under standard economics, but they could amplify cognitive mistakes made by others.[9]

Whether these are five separate biases or variations on one or two themes is less important here than the conclusion Sunstein draws from them:

> [P]redictable features of cognition will lead to a demand for regulation that is unlikely to be based on the facts. When people ask for regulation because of fears fueled by availability cascades, and when the benefits from the risk-producing activity are not registering, it would be highly desirable to create cost-benefit filters on their requests (Sunstein, 2000, p. 1072).

In other words, people will make mistakes as voters in their demand for regulation as voters. "Cost-benefit filters" will exclude their choices and substitute those justified by a separate analysis of the options at hand.

Sunstein has identified one important issue for BCA, which is to justify substituting technical expert analysis for what one would obtain through electoral processes. However, one need not invoke cognitive biases to justify this. One can appeal to, among other things, one person/one vote (as opposed to one dollar of surplus/one vote), the median voter rule, or Condorcet cycles. In addition, voters can choose only among candidates with bundles of positions on many issues rather than (in most cases) on whether and how to design a policy regarding a specific public project or health, safety, environmental, or welfare issue.

However, those failures do not support the general critique of democratic decision making that Sunstein makes to justify centralized decision-making, without regard to how those decisions get made (hence the scare quotes earlier):

> None of the cognitive points made here supports any particular understanding of cost-benefit analysis. Certainly I do not mean to embrace the controversial and indeed implausible proposition that all regulatory decisions should be made by aggregating private willingness to pay, as if economic efficiency is or should be the goal of all regulation (Sunstein 2000, p. 1060).

9 Imagine two new restaurants across the street from one another, and a succession of diners walking down the street. The first may have information that one restaurant is better, or may have none, and make a random choice. The second diner may have information that one is better, but if she sees the first diner in the other restaurant, will assign some likelihood that the first diner knew something as well. Or, if the second diner has no information, she may just decide to follow the first diner, in case that diner was informed. If a third diner sees two in the same restaurant, he may decide that the chance that his information (if any) outweighs theirs is so small as to make it virtually automatic to follow the first two. Any subsequent diner seeing three in one will follow them, etc. Hence, early information, even if incorrect, can lead to an outcome where everyone eats in one restaurant, even if their aggregate information indicates the other is better.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

Sunstein lists eight principles that an "incompletely-theorized" BCA could have that, in his view, would be supported by "a wide range of reasonable people" (Sunstein, 2000, p. 1092). The principles essentially posit identification of benefits and costs, quantification where possible with presumptive and adjustable ceilings and floors for mortality effects, recognition of qualitative effects, and at the end, a showing of "more good than harm, on a reasonable view about valuation of both benefits and costs" (Sunstein, 2000, p. 1095). "Social fear not based on evidence" justifies "education and reassurance rather than increased regulation" (Sunstein, 2000, p. 1095).

Notably, Sunstein did not find that markets are as susceptible to the failures that plague political participation. Specifically, he states:

> Some of these problems may infect market behavior as well, and when this is so there is a problem with using private willingness to pay as the basis for regulation ... [b]ut markets contain some safeguards against these errors, through the budget constraint and opportunities for learning (Sunstein, 2000, p. 1065, n. 14).

As we see below, he has apparently become less sanguine regarding whether to defer to choices that arise from market processes.

Whatever one thinks about the democratic implications of Sunstein's view, it does not offer a substitute for BCA based upon BE. It argues only that because BE implies that democratic decisions are flawed, BCA is justified, where what we mean by BCA remains to be decided.

4.2 Robert Sugden (2008)

Sugden's position is the opposite of Sunstein's. While BE supports the view that revealed benefit and cost data lead to wrong answers in policy contexts, it also supports the view that actions based on revealed preferences in market contexts will be similarly mistaken. Examples above include the apparent failure to make energy efficiency investments with high payoffs and that getting a pension plan seems unduly sensitive to whether one opts in or opts out. Another claim Sugden discussed is that consumers in a cafeteria line are unduly influenced by food placement, overbuying foods placed earlier in the line rather than what's healthier.

Taking a view explicitly against that espoused by Thaler and Sunstein in *Nudge* (Thaler & Sunstein, 2009), Sugden criticized the claim that these errors justify policy interventions, including "soft paternalism" or "libertarian paternalism" that preserves both options for the consumer. Examples include mandating

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

"opt-out" for pensions or placing healthier items ahead in cafeteria lines, leading people to choose better options, while preserving the ability of those who genuinely prefer the seemingly harmful option to get it. Sugden's most fundamental critique is against Thaler and Sunstein's view that anti-paternalism is impossible because choices are inevitably biased. Sugden responded that the choice is not between paternalistic standpoints but between intervention and Thaler and Sunstein's "surprising omission" (Sugden, 2008, p. 236) – maximizing profit, that is, letting entrepreneurs and the market sort out the opportunities. Underlying this is that the virtue of markets is not so much accuracy according to pre-existing preference but the facilitation of mutually preferred exchanges, if only at the time and context of the exchange. Moreover, Sugden's critique of Thaler and Sunstein's position show that Sunstein's (2000) earlier concerns with cognitive failures have gone beyond the political process to include market processes as well.

One can go further. The Sunstein-Thaler position is that default choice (e.g., opt-in or opt-out for pensions) is uninformative, therefore any effect it has on decisions must reflect cognitive limitations or biases. However, default options are not chosen randomly. If one asks why, in a competitive economy, particular options are selected, the likely answer is that the default option has been preferred more often more of the time than others. In the pension example, a new employee could reasonable infer that if getting (or not getting) the pension is the default option, it's because that's what most employees preferred, and given the cost of figuring out one's future economics position, may reasonably decide to go along. As Sunstein pointed out in the voting context, this can lead to an erroneous (if rational) outcome because of information cascades – early errors can bias the system (Sunstein, 2000, pp. 1066–1067).[10] However, changing the default may be throwing away information; that too is not neutral.

Whether Sugden's critique can save BCA from BE is a different matter. Sugden (2005) suggests that one can still use BCA when the measures of consumer surplus are based on simulating what would happen were there a market in place. However, the justification for BCA is a belief that there is a market failure justifying a policy intervention. Although one can justify BCA on standard economic grounds by saying that it is intended to replicate the outcome of a market that was not failing, this is essentially an efficiency argument. Sugden's defense of markets on the grounds that it reflects exchanges made between consenting parties does not justify BCA; the "consent" in BCA is hypothetical, not reflecting

10 Farrell and Saloner (1985) similar argued that inefficient technological standards can end up being adopted because early choices based on relatively minor considerations make changes down the road difficult. Liebowitz and Margolis (1995) critique the empirical relevance of this theoretical possibility and the inefficiency of "path dependence" more broadly.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

an actual exchange between the winners and losers (Dworkin, 1980). Sugden (2008, pp. 237–238) said, "I speculate that Sunstein and Thaler have convinced themselves that paternalism is inevitable by thinking of the anti-paternalist position as a particular kind of recommendation *to a planner*" (emphasis in original). If BE is empirically plausible, the challenges to BCA remain.

4.3 Douglas Bernheim and Antonio Rangel (2009)

Bernheim and Rangel (B-R) address the behavioral economics challenge through choice-theoretic formal modeling. They describe choice C as a function of an ordered pair (X, d), where X is a set of options and d the "ancillary condition" that may affect the selection of out of X. Formally, $C(X, d)\ \varepsilon\ X$. Under standard economics, the choice $C(X, d)$ is independent of d; under BE, d can matter in that there are some ancillary conditions d' and d'' such that $C(X, d') \neq C(X, d'')$.[11]

The definitional challenge is being able to identify when something is an ancillary condition and when it defines the objects from which one is choosing, that is, when the condition is uninformative or informative regarding properties of interest about X to the chooser. B-R claim to have a framework for policy analysis "no matter how one draws the line" between choice objects and ancillary conditions. They are able to define a preference function including ancillary conditions, but only with the implication that that there are no "cycles" in which persons can be given a sequence of binary choices under different ancillary conditions such that they end up where they started. This is a very strong property in the BE context, since it rules out preference reversals that give BE its empirical authority, that is, depending how a question is framed, one might choose x' over x'' or x'' over x' out of a choice set.

With regard to cost benefit analysis, B-R can only say that when ancillary conditions matter, measures of compensating variation will span some range. The minimum is the smallest amount that guarantees that y with that compensation would be preferred to x under all ancillary decision conditions, and the maximum is the largest amount that would keep x preferred to y despite the compensation. If decision contexts do not matter, these amounts will be the same. If they do matter, B-R say "the welfare effect of a change in the ancillary condition, by itself, is always ambiguous" and "the compensating variations … associated

11 Sugden (2008, pp. 241–242) characterizes different, inconsistent choices with different ancillary conditions as a "multiple selves" problem. Jon Elster (1979, 1985) has invoked "multiple selves" arguments to justify self-paternalism – using policies to precommit to restrict future choices to prevent actions the present self views as harmful – and to explain weakness of will.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

with each step in a sequence of changes needn't be additive" (Bernheim & Rangel, 2009, p. 73).

B-R set some limits on the potential range of surplus measures, but it still requires important restrictions to get even a limited preference ordering for which such a range can be defined. Under that framework, policy makers still have to decide when a "generalized choice situation" is sufficiently suspect to be disregarded. In other words, the policy maker has to decide which choices are legitimate and which are not. They hold out some hope that one could empirically determine when a choice is anomalous, based on when it appears inconsistent with other choices or is otherwise unusual. On the other hand, they note that if one thinks that decision contexts convey information actually relevant to a choice, one cannot determine that something is inconsistent – the ancillary conditions were not just ancillary. B-R also observed that a rare ancillary condition might be the one that gets people to give the right answer.

5 What can replace BCA?

If behavioral economics strips benefit cost analysis of its methodological justifications, and appeals to voting biases, market analogies, and generalized preference models do not solve the problem, we need to consider what might replace it. After identifying five alternatives and their shortcomings – some of which have been reviewed above – I conclude that a sixth imperfect alternative may be our best bet.

1. *Identifying the true willingness to pay.* A first option is for the policy maker to figure out the true WTP, using sources other than market data. In a sense, they do that when applying stated preference or contingent valuation surveys to elicit willingness to pay in the absence of market data (Carson & Hanneman, 2005). This has the advantage of relying on decentralized data. However, these methods are themselves fraught with biases in framing and context that lead to erroneous or inconsistent results. Also, because survey responses are inevitably not constrained by actual payments, survey responses often have failed to match market choices. If market and survey data are unreliable, one may be left with having the analyst decide which WTP estimates are reliable how to correct others.
2. *Libertarian paternalism.* Criticisms of the libertarian paternalism argument, that one can adjust the default option to induce the outcome most consumers genuinely want, have been presented above. In addition, one should recognize that few if any contexts in which BCA would be employed involve mere

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

nudges. BCA is employed primarily to look at non-marginal public investments in transportation, health, safety, environmental protection, education, research, and to address numerous other policy objectives. In those contexts, one does not have the ability to let those who want to opt out of the policy to do so. Even if libertarian paternalism could overcome the arguments against it in market contexts, it offers little in the way of an alternative to BCA when BCA is important.

3. *Using post intervention rather than pre intervention WTP.* A specific tack could be to make a policy intervention based on the idea that people would be better off afterwards even though they believe otherwise beforehand, and use the new WTP to justify the policy. Presumably, experience with the benefits will overcome biases that impede action beforehand. To put it another way, if people feel worse off after a policy intervention to correct their mistakes, it would be difficult to argue that they were mistaken regarding the effect of the policy before it was put into place.

   Brennan (2010) modeled such an approach in examining policies to evaluate energy efficiency subsidies calculated to get people to use high efficiency equipment. After having done so, the evaluation counts the energy savings as a benefit (without cost) that they were insufficiently appreciating prior to the subsidy. Brennan does not defend this approach; he notes it only to see how standard practices for evaluation of energy efficiency programs might be justified even though consumers fail to adopt these supposedly beneficial programs on their own. The problem with such an approach is apparent: How does one know what the post-policy WTP will be, if pre-policy WTP cannot be trusted?

4. *Changing underlying preferences.* Related to BE is the idea that contexts do not distort revealed preferences from true ones so much as change preferences themselves. Preference change presents numerous challenges to policy evaluation, if one is to retain the BCA principle that policies should be based on individual preferences. For example, if one wants to reduce emissions through preference change, should the target be based on the original or changed preferences. Increasing willingness to pay for "green" goods to reach one target will also change the target. Going the other way, one way to deal with environmental externalities would be change preferences so people no longer have an interest in a clean environment. The abhorrence of that option suggests that preference change means finding the "right" preferences, leaving open the issue of who gets to decide what the "right" preferences are (Brennan, 2006).

5. *Choosing policies democratically.* A last option here would be to have policies determined through democratic processes rather than through analytical

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

processes such as BCA. Electoral processes can satisfy most of the methodological advantages of BCA, especially decentralization and connection to benefit. However, as Sunstein argued, in a world where BE considerations matter, electoral processes themselves may be suspect. And as noted above, even without BE considerations, electoral processes are unlikely to be sufficiently specific to allow voters to speak to each policy for which BCA is relevant.

6 Democratic delegation and suggested principles

With flaws in all these other methods, one possibility remains: democratic delegation. People can choose through voting, accompanied perhaps by collective deliberation (Fishkin, 2009), the representatives who will make decisions for the voters that they believe they would incorrectly make for themselves. They would choose the representatives based on the criteria the representatives would use to make those decisions and address redistributive consequences. This reproduces some of the advantages of BCA, albeit indirectly. The process presumably is tied to benefit. It rests on decentralization to the extent that the public as a whole determines how policy will be adjusted in light of foreseeable cognitive biases and limitations. It will be less adept at coming up with a measure of benefits and costs consistent with judging the net effects of a policy, and suggesting how winners could compensate losers.

Two important considerations, one descriptive and one normative, lend some support to democratic delegation. The descriptive consideration is that democracies already do this. Beyond familiar policy processes, a few examples may illustrate how delegation works to solve problems based on bounded individual decision capacity. Price regulation of utility services has a clear standard rationale as a means to control monopoly market power, but it also may be a way for consumers to delegate choices regarding price and quality to central authorities with expertise, where search costs outweigh likely gains from picking a differentiated product meeting idiosyncratic desires (Brennan, 2007). Another example of responses to imperfect choices could be nominally paternalistic laws that may be a response to individuals' preferences to restrict their own actions in order to avoid bad outcomes they might fall into (Elster, 1979; Schelling, 1984). An example of this sort of self-paternalism could be drug laws, where the point is not for me to prevent you from taking drugs, but for me to prevent myself from taking them.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

On the normative side, democratic delegation is necessary even when BCA is accurate on its own terms, because efficiency is not everything. A policy can create net benefits so winners could compensate losers, but the policy process still has to decide, by commission or default, how the benefits and costs will be distributed. Apart from efficiency and distribution, some policies may affect fundamental rights to goods and services necessary to participate as a citizen and to pursue one's life objectives, such as freedom of speech and access to basic education.[12] The importance of policy considerations other than efficiency means that a process with the mechanical (if advantageously so) properties of BCA will not suffice much of the time.[13]

Still, the hope is that one would be able to do more than to kick the can to the voters. This is especially so because in practice such processes are likely to suffer from biases due to unbalanced interest group influence (Olson, 1971). Moreover, decision makers can be subject to the same BE-identified biases and limitations as do consumers and voters, to the potential detriment of policy outcomes (Cooper & Kovacic, 2012). Ideally, we would want to provide principles that these representative delegates should offer the public if one is making choices electorally.

Three potential principles come to mind. First, following Sugden (2005), perhaps democratically delegated decision makers should continue to use BCA because it simulates the outcomes of markets. His concept of market simulation rests on two premises. He observed that markets respond to revealed willingness to pay regardless of whether that willingness maximizes the individual's welfare by any separate criterion, and one need not expect policy to do differently. He also suggested that, to the extent that revealed willingness to pay is incoherent, in that it varies according to how a policy is framed – as Bernheim and Winston describe, for example – one should attempt to estimate willingness to pay by framing the choice as the policy would actually be implemented. Sugden proposed that policy evaluators should use the estimate that matches the funding choice, what he calls the "actual finance principle" (Sugden, 2005, pp. 140–141).[14]

12 This is obviously a topic far beyond the scope of this paper. We have found Rawls (1971) a useful starting point.

13 However, it will suffice a lot of the time, since many if not most of the projects for which a BCA is called for do probably do not confront fundamental rights or have society-wide distributive consequences.

14 Sugden (2005, p. 154) applies this to the familiar difference between how much people are willing to pay (WTP) or an environmental benefit and how much they would have to be paid to accept its disappearance ("willingness to accept" or WTA). If, as is normally the case, a policy would involve people paying for the environmental benefit, either as taxpayers or consumers from suppliers subject to regulation to provide that benefit, then one should base benefits on WTP rather than WTA.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

He justifies this market simulation test as it best replicates the surplus that would accrue under actual transactions (Sugden, 2005, p. 138).

However, as noted earlier, to the extent behavioral economics is empirically validated, the link between actual preference and revealed preference remains at least somewhat broken. A complement to recommending BCA to these democratically delegated decision makers is to identify criteria for when behavioral economics should be given weight and when it should not. Along those lines, we propose the following: Suppose behavioral economics identifies an action as irrational, mistaken, or a cognition failure for some person. If that person continues to do that seemingly irrational action after being informed of its putative irrationality, behavioral economics no longer applies, and conventional economics becomes the appropriate method for explanation and evaluation.[15,16] Judging the relevance of this criterion can allow these decision makers to decide when it is permissible after all to apply BCA.[17] Moreover, it provides added justification for addressing potential difficulties in applying BCA through information or education programs.

This leads to a third recommendation. While political deliberation continues in democratic processes that lead to the delegation, one should require very strong evidence that revealed preferences are inaccurate guides for benefit and costs. Because benefit-cost analysis has the important methodological advantages identified here, decisions to abandon it ought not be taken lightly.

7 Conclusion

Apart from its particular ties to economics, benefit-cost analysis has a number of important methodological advantages as a decentralized, interpersonal

15 This is not the same as the related empirical finding that in market settings, revealed preferences are less susceptible to the errors identified in behavioral economics (List, 2006). The similarity is that the explanation for findings such these is that people learn through repeated interactions (see also Sunstein's pre-*Nudge* standpoint discussed above). This suggestion broadens how learning may occur, from repeated instruction to a single explanation.

16 Let us invoke a personal example. I have a number of guitars, a few of which would command a high price in the "vintage" guitar market. I could sell some of these for more than I would be willing to pay for them. This appears irrational, because in foregoing the opportunity to sell them, I am in effect paying a price I claim I would not pay. However, as I know this, it is difficult to contend that this is irrational. Rather, there must be something else going on, e.g., uncertainty about trusting unknown buyers or the real option value of keeping the guitars because I might have a hard time finding a model of that vintage if I want one later on.

17 The need for judgment may arise in part because advocates of the behavioral economic perspective could argue that the person failed to understand the information and explanations.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

and internally consistent method for assessing benefits and identifying Pareto improvements. In claiming that revealed willingness to pay differs from actual value, behavioral economics challenges all of these advantages. Cass Sunstein, Robert Sugden, and Douglas Bernheim and Antonio Rangel each offer potential ways to reconcile behavioral economics with cost-benefit analysis, but for a variety of reasons, all fail to do so.

After reviewing a number of alternatives to benefit-cost analysis, democratic delegation of authority to make judgments regarding the net benefits of regulation seems the least bad. It is not unfamiliar, as democratically delegated regulators also have to make decisions on matters outside the scope of BCA, such as distribution of the benefits and costs. We offer three guiding principles: (from Sugden, 2005) consider the merits of viewing policy evaluation in market simulation rather than welfare terms, restrict behavioral economics to settings where customers have not learned or been informed of the putative error of their decisions, and demand clear evidence of differences between revealed and actual willingness-to-pay before rejecting the findings of conventional benefit-cost analysis.

References

Bernheim, D., & Rangel, A. (2009). Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics, 124*, 51–104.

Brennan, T. (2006). 'Green' preferences as policy instrument. *Ecological Economics, 56*, 144–154.

Brennan, T. (2007). Consumer preference not to choose: methodological and policy implications. *Energy Policy, 35*, 1616–1627.

Brennan, T. (2010). Optimal energy efficiency policies and regulatory demand-side management tests: how well do they match? *Energy Policy, 38*, 3874–3885.

Carson, R., & Hanneman, M. (2005). Contingent valuation. In: K. -G. Mäler & J. R. Vincent (Eds.), *Handbook of Environmental Economics* 2 (pp. 821–935). Amsterdam: Elsevier.

Choi, J., Laibson, D., Madrian, B., & Metrick, A. (2002). defined contribution pensions: plan rules, participant decisions, and the path of least resistance. In: J. Poterba (Ed.), *Tax Policy and the Economy,* 2 (pp. 67–114). Cambridge, MA: MIT Press.

Cooper, J., & Kovacic, W. (2012). Behavioral economics: implications for regulatory behavior. *Journal of Regulatory Economics, 41*, 41–58.

Dworkin, R. (1980). Why efficiency? *Hofstra Law Review, 8*, 563–590.

Elster, J. (1979). *Ulysses and the sirens.* Cambridge: Cambridge University Press.

Elster, J. (1985). Weakness of will and the free rider problem. *Economics and Philosophy, 1*, 231–265.

Enkvist, P. -A., Nauclér, T., & Rosander, J. (2007). A cost curve for greenhouse gas reduction. *The McKinsey Quarterly, 1*, 35–45.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *Rand Journal of Economics, 16*, 70–83.
Fishkin, J. (2009). *When the people speak: deliberative democracy and public consultation*. New York: Oxford University Press.
Horowitz, J., & McConnell, K. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management, 44*, 426–447.
Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: the endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*, 193–206.
Knetsch, J. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review, 79*, 1277–1284.
Liebowitz, S., & Margolis, S. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics and Organization, 11*, 205–226.
List, J. (2006). The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy, 114*, 1–37.
Mullainathan, S., & Thaler, R. (2001). Behavioral economics. In: N. Smelser and P. Baltes (Eds.). *International Encyclopedia of the Social and Behavioral Sciences,* (pp. 1094–1100). Amsterdam: Elsevier.
Olson, M. (1971). *The logic of collective action: public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Revesz, R., & Livermore, M. (2008). *Retaking rationality: how cost-benefit analysis can better protect the environment and our health*. Oxford: Oxford University Press.
Schelling, T. (1984). Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review Papers and Proceedings, 74*, 1–11.
Sugden, R. (2005). Coping with preference anomalies in cost-benefit analysis: a market simulation approach. *Environmental and Resource Economics, 32*, 129–160.
Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy, 19*, 226–248.
Sunstein, C. (2000). Cognition and cost-benefit analysis. *Journal of Legal Studies, 29*, 1059–1103.
Thaler, R., & Sunstein, C. (2003). Libertarian paternalism. *American Economic Review Papers and Proceedings, 93*, 175–179.
Thaler, R., & Sunstein, C. (2009). *Nudge: improving decisions about health, wealth, and happiness*. New York: Penguin.
Young, H. P. (1993). *Equity: in theory and practice*. Princeton: Princeton University Press.

https://doi.org/10.1515/jbca-2013-0022 Published online by Cambridge University Press