University of East Anglia

Does behavioural public policy correct individuals' 'biases' and 'errors'? Or is it just paternalism?

Robert Sugden

Talk at European Commission Behavioral Insights seminar series

7 November 2023

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 670103.

My starting point: the problem of reconciling normative economics with the findings of behavioural economics.

From the 1940s to the 1990s, there was a consensus about how to do normative economics: **neoclassical welfare economics**.

This was the basis for economics of market regulation, and for cost-benefit analysis.

Neoclassical welfare economics uses preference satisfaction as its normative criterion.

Presupposition: individuals have **integrated** preferences (i.e. well-defined, stable, context-independent) which are revealed in their choices.

But findings of behavioural economics show that individuals' choices often *don't* reveal integrated preferences.

Instead: predictable context-dependent effects that can be explained by psychological theories but seem not to correspond with good reasons for differences in behaviour. E.g. 'Apples and Mars bars' experiment…

Apples and Mars bars (experiment by Read and van Leeuwen, 1998):

Office workers are approached either just after lunch, or in late afternoon; they are offered a choice of snacks that will be delivered at a fixed time (just after lunch, or in late afternoon) *a week later*; they have to choose now which snack to have. Some snacks are 'healthy' (e.g. apple), some are 'unhealthy' (e.g. Mars bar).

Result: irrespective of the *delivery time*, individuals are more likely to choose unhealthy snacks if the *choice* is made in late afternoon.

Notice:

- The difference in choice has a psychological cause (people are hungrier in late afternoon; thoughts about hunger-satisfying properties of food are more salient to people who are hungry).

- The psychological difference doesn't seem to be a *good reason* for the difference in choice (the options are familiar and daily fluctuations in hunger are very predictable);

- but there is no obvious way of determining whether it is more rational to choose an apple or a Mars bar (at either time of day).

So, a problem for normative economics, which began to be taken seriously from about 2000.

If (revealed) preferences vary according to factors that are not relevant to individuals' interests or welfare, how can we justify using preference-satisfaction as a normative criterion?

And if we can't, what normative criterion *should* we use?

Emerging consensus among behavioural economists: **behavioural welfare economics.**

[Proposed by: Bleichrodt, Pinto-Prades and Wakker (2001); Sunstein and Thaler (2003, 2008); Camerer, Issacharoff, Loewenstein, O'Donoghue and Rabin (2003); Kőszegi and Rabin (2007); Salant and Rubinstein (2008); (with qualifications) Bernheim and Rangel (2009).

Approved by Kahneman (personal communication).

Characterisation and critique of behavioural welfare economics: Infante, Lecouteux and Sugden, *Journal of Economic Methodology* 2016.]

Contrasting approach proposed by Sugden (2004) but hasn't had many takers.

Behavioural welfare economics

I'll focus on Sunstein and Thaler, as the most prominent advocates of behavioural welfare economics …

Sunstein and Thaler (2003). 'Libertarian paternalism is not an oxymoron', *University of Chicago Law Review*.

Thaler and Sunstein (2008). *Nudge*.

Richard Thaler

Cass Sunstein

Sunstein and Thaler contrast individuals in traditional economics with real human beings ('Econs' vs 'Humans'). The findings of behavioural economics make the criterion of (revealed) preference-satisfaction 'incoherent', a 'non-starter'. They support this claim by using the *cafeteria* example…

Premise: customers' choices between food items depend on the positions in which they are displayed (i.e. are **context-dependent**). Knowing this, how should the cafeteria director choose the positions?

'[the customers] lack well-formed preferences, in the sense of preferences that are firmly held and preexist the director's own choices about how to order the relevant items [along the counter]. If the arrangement of the alternatives has a significant effect on the selections the customers make, then their true [revealed] 'preferences' do not formally exist.'

So what criterion should the director use? S&T say she should aim to

'make choosers better off, *as judged by themselves*' (2008, their italics).

Implication: the director/ planner/ choice architect should respect individuals' *subjective judgements* about their own welfare (NB: S&T are not invoking an objective concept of welfare)….

… but should *not* presuppose that these judgements are revealed in individuals' choices.

Problem: When choices are context-dependent, how are we to understand these judgements? And how is the planner to reconstruct them?

The nearest S&T get to answering these questions is in discussing *decision-making errors*…

Immediately after the remark about making choosers better off, as judged by themselves, S&T say they will show that:

'in many cases, individuals make pretty bad decisions – decisions that they would not have made if they had **paid full attention** and **possessed complete information, unlimited cognitive abilities, and complete self-control'**.

Such decisions are 'inferior decisions in terms of their [i.e. the individuals'] own welfare'.

Implication: S&T's welfare criterion is the satisfaction of the **latent preferences** [my term] that an individual would have revealed *in the absence of reasoning imperfections*. Behavioural welfare economics = reconstructing what the individual would have chosen in the absence of reasoning imperfections.

Context-dependence of **revealed** preferences is the problem S&T address (the nudge programme works only if revealed preferences are context-dependent).

But there is an implicit assumption that **latent preferences are context-independent.**

S&T never justify this assumption. It depends on a peculiar model of human agency…

The model of the inner rational agent (as critiqued by Infante et al.)

S&T's implicit model of human agency: a faulty Econ.

Inside the human being, there is a neoclassically rational agent, with integrated preferences. These are the latent preferences that can be reconstructed by behavioural welfare economics.

But **the rational agent is trapped in a psychological shell**. Its interactions with the world are mediated by the shell.

Properties of the psychological shell cause **errors** in decisions:

- lack of memory capacity, so the inner agent does not have access to all the information it requires;
- lack of attention, so relevant information does not always reach the inner agent;
- lack of cognitive ability, so computations that the inner agent requires are not always accurate;
- lack of self-control, so the inner agent's decisions are not always executed.

What's wrong with this model?

It doesn't take psychology seriously.

The starting point for behavioural economics was recognising that the mental processes actually used in decision-making do not necessarily generate choices with the rationality properties assumed in economics.

An obvious corollary, noted by Kahneman (1996): **rational choice is not self-explanatory**, i.e. behaviour that is consistent with the standard economic theory is just as much in need of psychological explanation as are 'anomalies'.

But in the model of the inner rational agent, human psychology is represented as a set of forces which affect behaviour by *interfering with* rational choice. Rational choice itself is not given any psychological explanation.

There is no psychological explanation of why latent preferences exist at all. If *actual* choices are determined by context-dependent cues (e.g. the Mars bar case), what is the function of latent preferences?

An example relevant to market regulation:

Point-of-sale offers of insurance or maintenance contracts for consumer durables (e.g. GAP* insurance for new cars).

*guaranteed asset protection

The facts: many consumers would (i) accept high-price point-of-sale offers but (ii) if these offers were not available at point of sale, later low-price offers would be rejected. This is a case of context-dependent preference. A plausible psychological explanation: different degrees of attention to insurance.

But this can be 'explained' by either of two opposite 'biases':

- over-attention to a temporarily salient stimulus at the point of sale (i.e. point-of-sale decisions are errors);
- under-attention after buying car (compare 'procrastination', failure to switch energy suppliers.

Empirical psychology identifies and explains context-dependent choice, but doesn't have a concept of *correct* choice or *correct* distribution of attention – and doesn't need one. In this context, **'bias' and 'error' are not empirical concepts.**

How does behavioural welfare economics produce concrete policy recommendations?

Typically, by appealing to common-sense judgements about welfare, e.g. justifying nudges towards healthy lifestyles by referring to familiar statistics about (e.g.) alcohol consumption and health:

'With respect to diet, smoking, and drinking, people's current choices cannot reasonably be claimed to be the best means of promoting their well-being.' (S&T, *Nudge*)

Then by inferring that a heavy drinker must be making errors of attention, information, cognition or self-control.

Implicit assumption: each individual can articulate 'justifications' of his choices as welfare-maximising, and the behavioural economist can check these for errors.

But this is a rational-choice theorist's view of human decision-making. Buying a drink needn't be a proposition about your welfare, it can be just a response to the cues of the moment.

So, how should we reconcile normative and behavioural economics?

I have two proposals, one for market regulators, one for economic policy-making by democratic governments.

Proposal 1: The role of market regulators

Much the same as in neoclassical economics:

- promoting the effective operation of competitive markets, i.e. their capacity to provide a wide range of opportunities for mutually beneficial transactions;

- defining and enforcing standards of **'transactional fairness'** (roughly: **no deception** (firms' offers must be transparent) and **no hindrance** (of potential customers' attempts to access other firms' offers) [Lyons and Sugden, 'Transactional fairness in consumer markets', *Behavioural Public Policy*];

- but not trying to discover and satisfy consumers' 'true' preferences.

[The GAP insurance example: there is a failure of competition. The same good is being sold at two very different prices because of temporary monopoly power at the point of sale. This diagnosis is independent of whether buying GAP insurance is 'rational' or 'irrational'.]

Proposal 2: economic policy making by democratic governments

A democratic government is entitled to be paternalistic if it has an electoral mandate for this, but it should be honest about its reasons.

And more to the point: behavioural economists should be honest about the justifications for their policy proposals.

We shouldn't claim that a policy 'makes choosers better off, *as judged by themselves*' when what we mean is that it makes choosers better off, *as judged by us.*

Thank you for listening.