

## **‘On the Econ within’: a reply to Daniel Hausman**

Gerardo Infante (School of Economics, University of East Anglia)

Guilhem Lecouteux (Département d’économie, École Polytechnique)

Robert Sugden (School of Economics, University of East Anglia)

Our paper is a critique of an approach to normative economics – ‘behavioural welfare economics’, or BWE for short – which is advocated by many prominent behavioural economists. Daniel Hausman features in our argument as (in his words) an alleged philosophical fellow-traveller. Having written the paper in the hope of sparking off a debate about how behavioural economics should deal with normative questions, we are pleased that he has responded with a defence of some of the features of BWE that we criticised.

If we have read Hausman’s comment correctly, he does not disagree with our characterisation of what BWE is, or with our attribution of this approach to behavioural economists. In relation to our critique, the main features of BWE can be summarised as follows: (1) The approach is intended to apply to cases in which individuals’ revealed preferences depend on contextual factors that have little or no apparent relevance to those individuals’ interests or well-being. (2) The normative criterion is the satisfaction of each individual’s *latent* (or ‘purified’) preferences, defined as the preferences she would reveal in the absence of any errors that might be caused by limitations of attention, information, cognitive ability or self-control. (3) Latent preferences are interpreted as expressing individuals’ *subjective judgements* about their interests or well-being; they do not necessarily track objective properties of the external world (such as an individual’s monetary wealth or health status) or properties of passive experience (such as happiness in the hedonic sense). (4) In the cases to which BWE is to be applied, latent preferences are assumed to be context-independent.

In saying that BWE implicitly uses a model of an inner rational agent, we mean no more than that it has properties (1) to (4). Of course, behavioural economists do not maintain that a human being is *really* made up of a neoclassically rational entity encased in an error-prone psychological shell. But we think it is both true and illuminating that BWE proceeds *as if* human beings were like that. BWE invokes a model of rational agency when it assumes that, in the absence of error, an individual’s choices would reveal coherent preferences. This

rationality is modelled as an ‘inner’ property of the individual in the sense that, although rationality is not always revealed in actual behaviour, it is a continuing property of the individual’s capacity to form context-independent subjective judgements on the basis of error-free reasoning. The inner rational agent of BWE is no more and no less than that assumed capacity.

To clear the ground, we begin with the issue of whether Hausman is a fellow-traveller of BWE. We agree (and tried to make clear in the original paper) that he does not give unqualified endorsement to the approach characterised by properties (1) to (4). In particular, he does not treat the satisfaction of preferences (whether actual or latent) as a decisive criterion of well-being. For Hausman, an individual’s preferences express her subjective evaluative judgements about alternative outcomes. These judgements typically provide useful *information* about the determinants of the individual’s well-being, but preference-satisfaction does not *constitute* well-being. For example, the individual’s preferences might reflect a concern for other people’s well-being; that concern might be revealed in her choices without other people’s benefits being components of her own well-being. However, although Hausman sometimes talks about ‘objective’ well-being, we take him to be associating an individual’s well-being with the achievement of her ‘most serious objectives’ and of ‘what matters most’ to her, at least when these objectives are self-interested (p. 4). If this reading is right, Hausman’s concept of well-being, like that of latent preference in BWE, is linked to subjective judgement.

Hausman also differs from many advocates of BWE in his caution about the range of applicability of the concept of latent preference. However, as is implicit in his discussion of the cafeteria case, and as we document more fully in our paper, he endorses various claims about well-being that are prominent in the literature of BWE and which, in that literature, are grounded on claims about latent preferences. In these specific but paradigmatic cases, Hausman seems to accept that latent preferences are identifiable, context-independent, and a source of useful information about individuals’ judgements about what matters most to them (see, for example, the final sentence of his footnote 2). Our critique of the preference purification approach applies to Hausman’s position on these cases just as much as it does to orthodox BWE.

We now turn to the central issue: Is that critique sound? Hausman recognises the ‘irony’ in behavioural economists’ appeal to inner rational agents, but he finds our arguments unclear or unconvincing except insofar as they reduce to the familiar ‘epistemological’ and

‘normative’ objections to BWE (as presented, for example, by Hausman and Welch, 2010). Nevertheless, our paper is *not* a reformulation of those objections. The epistemological objection, as stated by Hausman, accepts that, in principle, it may be possible to define and observe context-independent latent preferences; the objection is that the elicitation of those preferences may be difficult in practice (pp. 4–6). The normative objection is that, if public policy is directed at the satisfaction of latent preferences, individual autonomy may not be properly respected. Our argument, in contrast, leads to a more fundamental conclusion – that there is no general reason to expect that context-independent latent preferences, as understood in BWE, exist at all.

Our argument is based on three linked criticisms of BWE. The first is that the concepts of latent preference and error lack adequate psychological foundations: in the best available psychological theories of context-dependent choice, those concepts play no explanatory role. For example, to explain why a person’s choices between multi-attribute consumer products vary according to the relative prominence with which the products are displayed, it is sufficient to explain how the weight given to each attribute in the deliberative process is influenced by the amount of attention that is given to it, and hence by whatever factors influence the distribution of attention. Nothing is added by labelling some particular distribution of attention, and the choices that this distribution induces, as ‘correct’, and the others as ‘error’. The second criticism is that, if BWE is to work as it is supposed to do, it needs an operational definition of error. BWE defines error in terms of specific impairments of ‘correct reasoning’, but gives no description of what this reasoning actually is. We are asked to believe that this reasoning is capable of generating context-independent preferences, but are not told how these preferences are generated. The third criticism is that, on a coherent and plausible account of reasoning, correct reasoning may sometimes fail to determine any uniquely correct preference ranking of choice options. In such cases, one is not entitled to infer errors of reasoning from observations of context-dependent choice – even if the relevant contextual factors clearly have no bearing on the decision-maker’s interests or well-being.

It seems to us self-evident that these points, if correct, are fundamental methodological criticisms of BWE – and all the more so because of how the research programme of behavioural economics is understood by its practitioners. Behavioural economics surely *is* economics with psychological foundations. One of the main ways in which behavioural economics differentiates itself from neoclassical economics is in not accepting *a priori* assumptions about rational agency, unsupported by empirical evidence.

(Recall Sunstein and Thaler on the contrast between ‘Humans’ and ‘Econs’.) It is more than ironic if BWE has to explain observed deviations from neoclassical predictions by invoking ungrounded and untested assumptions about latent rationality.

Most advocates of BWE are surprisingly unconcerned about whether claims about latent preferences can be grounded on general theoretical principles. There is a prevailing tendency to brush aside awkward questions with bland assertions to the effect that, in whatever case is currently under discussion, the assumed properties of latent preferences are so obviously true that it would be pedantic to ask for rigorous justifications. (See, for example, Sugden (2009), who documents this tendency in Sunstein and Thaler’s justifications of nudging.) It seems to us that Hausman is following a similar argumentative strategy. Revealingly, he does not disagree with our specific criticisms of Bleichrodt, Pinto-Prades and Wakker’s attempt – one of the very few serious attempts to date – to develop a general and theoretically-grounded method for defining and eliciting latent preferences; he simply rejects that method as ‘question-begging’. But his two positive suggestions about how to solve the problem addressed by Bleichrodt et al. are at least as vulnerable to that charge.

One of these suggestions is to run experiments which elicit individuals’ preferences in decision environments that present alternatives ‘without bias’ (p. 5). But decades of experimental investigation of preference elicitation have revealed that the concept of an ‘unbiased’ decision environment is riddled with problems. To take just one such problem – one that Hausman (1992, pp. 227–244) has written about in some detail – consider preference reversal. If we want to discover which of two options an individual prefers, we might offer her a straight choice between the two and then infer a preference for the chosen option. Alternatively, we might ask her to consider each option separately and in each case to say what amount of money she finds exactly as preferable, and then infer a preference for the more highly valued option. Neither of these methods is obviously biased, but for some kinds of option pairs they elicit systematically different preferences. Psychologically, the source of the problem seems to be that choice tasks and valuation tasks call into play different mental operations: this is another case of context-dependent choice. But how are we to determine which (if either) of these mental operations delivers true preferences and which does not?

Hausman’s other suggestion is contained in his discussion of the person (Joe in our paper) whose choices between cake and fruit are context-dependent (pp. 3–4). He describes a case in which, he claims, one can infer that Joe has a latent preference for fruit without postulating ‘the existence of an inner agent who is capable of weighing the various

considerations ... and determining an overall preference'. The relevant facts of the case are that Joe is 'quite generally concerned about his health and appearance and deeply regrets his occasional sugar binges', which Hausman interprets as evidence that 'health and appearance matter more to [Joe] than the pleasures of sweets'. But how is this argument supposed to work?

Let us flesh out this case a little more, editing out the loaded language of 'binges'. Imagine a Joe who consistently maintains that both health and appearance matter a lot to him. With similar consistency, he maintains that life's small pleasures, such as the pleasure of eating a sugary and creamy cake when hungry, matter too. He generally thinks of resoluteness as a virtue, but he believes that there is also a place in life for spontaneity: it is possible to be too slavishly constrained by one's previous resolutions. Joe's problem in the cafeteria is not to decide whether *in total* (whatever that may mean) health and appearance matter more than the enjoyment of small pleasures. It is to choose between *different combinations of* aspects of well-being – health, appearance, small pleasures – all of which matter to him. If he has a standing resolution to eat healthily, he also has to decide whether this is an occasion for resoluteness or for spontaneity. In other words: he has to weigh the various considerations and determine an overall preference.

The problem for normative economics is that this weighing operation is influenced by contextual cues. In our paper, we referred to Kahneman's hypothesis that people's judgements about the relative importance of different dimensions of their lives depend on what they are currently attending to. Suppose this is true of Joe, and suppose the cafeteria is one in which cake is displayed prominently. Joe-in-the-cafeteria strikes a balance between the relevant aspects of well-being and between the virtues of resoluteness and spontaneity, and the balance comes down in favour of the cake. But in other contexts, he strikes a different balance. Think of Joe-in-front-of-the-mirror, not currently feeling hungry but attending to his expanding waistline. *Now* he deeply regrets the cake-buying decisions he has made in the cafeteria. But it might be equally true to say that, had Joe-in-the-cafeteria discovered that some previous decision (say about how much money to bring with him) had prevented him from buying the cake, he would *then* have regretted having to eat the cheaper fruit.

If Hausman's argument is to apply to this fleshed-out case, we have to find it obvious that Joe-in-the-cafeteria is reasoning erroneously, and that Joe-in-front-of-the-mirror is the author of Joe's latent preferences. Even if this *were* obvious, it would still be the case that

those preferences were the product of a process of weighing various considerations and determining an overall preference, and (to support the claim that Joe's choice of cake was a mistake) one would have to attribute to Joe-in-the-cafeteria a latent capacity to work through this process correctly. In other words, one would have to use the model of an inner rational agent.

But, in the absence of a specification of what correct reasoning is, what is the truth condition for the supposedly obvious proposition that Joe's latent preference is for fruit rather than cake? Hausman says that this proposition is grounded in features of Joe's psychology which allow us to attribute 'particular objectives and flaws' to him. But the crucial problem, consistently evaded by BWE, is to determine in a scientifically defensible way *which* features of human psychology represent latent objectives and which features can be classified as reasoning flaws. Hausman adds that his conclusions are derived from 'an account, which the agent can affirm or dispute, of what matters most to this flesh-and-blood individual' (p. 4). But who or what is 'the agent'? In the story as we have told it, Joe-in-front-of-the-mirror weights the relevant aspects of well-being in a way that supports a preference for fruit; Joe-in-the-cafeteria weights them in a way that supports a preference for cake. If 'the agent' is Joe the real human being, what he affirms about his well-being depends on the context in which he is deliberating, and so we are no further forward. Perhaps (as one distinguished advocate of BWE has suggested to us privately), we should ask which context Joe himself would choose as most representative of his true self. But then we face the same problem again: if Joe's ideas about the relative importance of different aspects of life are context-dependent, his conception of his true self is likely to be context-dependent too. It seems that what Hausman means by 'the agent' is some capacity for context-independent reasoning about well-being that is assumed to be latent in Joe – or, in other words, Joe's inner rational agent.

## References

- Hausman, Daniel (1992). *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hausman, Daniel and Brynn Welch (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy* 18: 123–136.

Sugden, Robert (2009). On nudging. *International Journal of the Economics of Business*,  
16: 365–373.