Behavioural Public Policy (2023), **7**, 212–218
doi:10.1017/bpp.2021.19

CAMBRIDGE
UNIVERSITY PRESS

RESPONSE

'Come on, man!' On errors, choice, and Hayekian behavioral economics

Cass R. Sunstein

Robert Walmsley University Professor, Harvard University, Cambridge, MA, USA
E-mail: csunstei@law.harvard.edu

(Received 27 April 2021; accepted 27 April 2021; first published online 15 September 2021)

Abstract

With respect to the views of dead thinkers, answers to many particular questions are often interpretive in Ronald Dworkin's sense. Such answers must attempt (1) to fit the materials to be interpreted and (2) to justify them, that is, to put them in the best constructive light. What looks like (1), or what purports to be (1), is often (2). That is, when a follower of Kant urges that 'Kant would say x', or that 'Kantianism entails y', the goal is to make the best constructive sense of Kant and Kantianism, not merely to adhere to something that Kant actually said. An approach to behavioral economics cannot claim to be Hayekian if it is rooted in enthusiasm for the abilities of planners to set prices and quantities, or if it sees the price system as a jumble of mistakes and errors. But within a not-so-narrow range, a variety of freedom-preserving approaches, alert to the epistemic limits of planners, can fairly claim to be Hayekian. Hayekian behavioral economics, I suggest, is an approach that (1) recognizes the importance and pervasiveness of individual errors, (2) emphasizes the epistemic limits of planners, (3) builds on individual choices rather than planner preferences, and (4) gives authority to choices made under epistemically favorable conditions, in which informational deficits and behavioral biases are least likely to be at work. The key step, of course, is (4). If it is properly elaborated, the resulting approach deserves respect. It is worthy of serious consideration, even if some of us, including the present author, would not entirely embrace it. In defending that proposition, the present essay responds to some critical remarks on behaviorally informed policy, including the resort to 'explainawaytions' (Matthew Rabin's term) for behavioral findings.

Keywords: behavioral economics; internalities; self-control; Hayek; knowledge problem

It is easy to imagine Benthamite behavioral economics; it would embrace Bentham's form of utilitarianism (and draw on Bentham's understanding of human motivation). It is not as easy, but not terribly hard, to imagine Aristotelian behavioral economics; it would be rooted in a conception of human flourishing (and draw on Aristotle's understanding of human motivation). Kantian behavioral economics would emphasize the importance of treating people as ends rather than means. It may or may not have a great deal to say on the positive side – compared with Bentham and Aristotle, Kant did not focus much on

people's actual motivations and behavior – but in terms of the moral foundations of behaviorally informed public policy, it would be highly prescriptive.

To be sure, general propositions do not decide concrete cases. Would Benthamite behavioral economics favor or disfavor a prohibition on the sale of cigarettes (Conly, 2013)? Would Aristotelian behavioral economics favor or disfavor taxes on sugar-sweetened beverages (Allcott *et al.*, 2019)? Would Kantian behavioral economics favor or disfavor automatic enrollment in savings policies? Would Bentham, Aristotle, or Kant embrace behavioral welfare economics as understood by Bernheim, with its choice-theoretic foundations (Bernheim, 2016)? You cannot obtain answers to these particular questions simply by reading and rereading the works of Bentham, Aristotle, and Kant. You cannot point to the categorical imperative (and thump the table). With respect to the views of dead thinkers, answers to many particular questions are inevitably *interpretive* in Ronald Dworkin's sense (Dworkin, 1985): they must attempt (1) to *fit* the materials to be interpreted and (2) to *justify* them, that is, to put them in the best constructive light.

It is, of course, true that if an approach does not fit the materials, it cannot claim to be interpretive. But with respect to many specific questions of interpretation, more than one conclusion will do well, or well enough, along the dimension of fit, and the real question is one of justification. What looks like (1), or what purports to be (1), is often (2). That is, when a follower of Kant urges that 'Kant would say x,' or that 'Kantianism entails y,' the enterprise may be to make the best constructive sense of Kant and Kantianism, not to follow some edict of Kant himself. The interpreter might even be said to be authoring a new chapter in a kind of chain novel, whose initial chapters were written by Kant, but which Kant did not complete (Dworkin, 1985).

Which understanding of Kant puts his views in the best constructive light? Because this is often the inevitable question, it should not be amazing to find that Jones' preferred interpretation of Kant fits with Jones' normative views, while Smith's interpretation fits with Smith's. The dispute between Jones and Smith might not be resolvable without speaking of *justification* – of what does, in fact, put Kant in the best constructive light. Inevitably, that judgment is an interpretation, in which we attempt to make the best constructive sense of what he actually said and thought.

My main goal in *Hayekian Behavioral Economics* (Sunstein, 2021) was interpretive in Dworkin's sense. I sought to sketch an approach that (1) recognizes the importance and pervasiveness of individual errors, (2) emphasizes the epistemic limits of planners, (3) builds on individual choices rather than planner preferences, and (4) gives authority to choices made under epistemically favorable conditions, in which informational deficits and behavioral biases are least likely to be at work. The heart of the article, and the key step, of course, is (4). On that step, there is a large and growing theoretical and empirical literature, with important contributions from Bernheim and Rangel (2007; 2009), Goldin (2015), Allcott and Knittel (2019), Allcott and Taubinsky (2015), Bernheim (2016), Bernheim and Taubinsky (2018), and others. One of my purposes was to unify that literature and to organize it under a general framework, which is insistently Hayekian in the sense that it is committed to (2) and (3), even as it is also committed to (1). To discipline (4), I pointed to five subsidiary questions:

https://doi.org/10.1017/bpp.2021.19 Published online by Cambridge University Press

(1) What do consistent choosers, unaffected by clearly irrelevant factors or 'frames', choose?
(2) What do informed choosers choose?
(3) What do active choosers choose? (If we focus on active choosers, we will protect against the possibility that outcomes are a product of inertia or procrastination.)
(4) In circumstances in which people are free of (say) present bias or unrealistic optimism, what do they choose?
(5) What do people choose when their viewscreen is broad, and they do not suffer from limited attention?

I also aimed to put Hayekian behavioral economics in an appealing light, though it does not, in fact, reflect my own view. I would not give decisive authority to individual choices, even under epistemically favorable conditions; human welfare is the criterion, not respect for choices, though the two are of course related (Sunstein, 2020a). If we respect choices, we will often and even generally promote welfare, but in important cases, respect for choices will undermine welfare, and welfare is authoritative (Conly, 2013; Sunstein, 2020a). The task of specifying 'welfare' is, of course, contentious and takes us into contested philosophical waters. Are we speaking in utilitarian terms? Bentham's version, or Mill's? Might we favor a nonutilitarian form of consequentialism? Might consequentialism import deontological considerations? The best answers to these questions will, I think, lead us to question Hayekian behavioral economics, as I have elaborated it, and to undertake the specification, or at least to notice the need for it (Sunstein, 2020b).

Robert Sugden (2021) suggests that my article is 'merely re-naming a familiar approach to normative economics', which, in his account, was 'initiated in' an article introduced by Richard Thaler and me in 2003. Nothing could be further from the truth. With the exception of wine (he likes it, I hate it), Thaler and I tend to agree, and I would not want to saddle Thaler with a view that I myself do not hold. I can report with some authority that the view we sketched in 2003 (which I still do hold, mostly), is not at all the same as the Hayekian view I sketched in 2021 (which, to repeat, is not my own). In 2003, Thaler and I argued for 'libertarian paternalism', that is, an approach that preserves freedom of choice while steering people in certain directions, but we did not elaborate (4) in any way. By emphasizing certain choice-attuned ways of deciding on the right defaults (Sunstein, 2020b), I suppose that we made some gestures in broadly Hayekian directions. But we did not engage Hayek's work in any terms, and we did not have a time-travel machine. Certainly, we could not have anticipated the important research and findings of Goldin, Allcott, Bernheim, Daubinsky, and others.

Rejecting my account of Hayekian behavioral economics, Sugden urges that the concepts of 'bias' and 'error' make sense only if 'individuals have context-independent latent preferences', which, in his view, they lack. I have no idea what the term 'latent preferences' means. (I have never used it, and I can't recall anyone else who has.) When information is absent and when behavioral biases are at work, things usually aren't all that complicated. We ought not to get all tied up in nouns and abstractions.

https://doi.org/10.1017/bpp.2021.19 Published online by Cambridge University Press

Suppose that I buy an apple thinking that it is an orange. We do not have to get fancy, or to speak of 'latent preferences', in order to say that I have made an error (Bernheim, 2016). Or suppose that someone buys a new car, believing that it has certain safety features; suppose that he would not buy the car if he did not hold that belief. Suppose that the car lacks those features. We do not have to get fancy, or to speak of 'latent preferences', in order to say that the purchaser has made an error. Or suppose that someone refuses to get vaccinated against COVID-19 because she is unrealistically optimistic (she says that 'she never gets sick') or because of availability bias (she read about someone who died as a result of getting vaccinated). We do not need to get fancy, or to speak of 'latent preferences', in order to conclude that a bias has led to an error (Allcott & Sunstein, 2015). If someone says that she will finish a project on February 1 and does not finish until April 15, overconfidence, and the planning fallacy, may well have led to the mistake (Kahneman *et al.*, 2021). To reach this conclusion, we do not need to say a word about 'latent preferences'.

Maybe you don't love those examples. Matthew Rabin has coined the term 'explainawaytions' for efforts to explain away behavioral findings by offering some not-certifiably-crazy rational explanation, and then declaring victory. I am acutely aware that if someone refuses to get vaccinated, drinks a ton, gets addicted to opioids, or eats enough to become morbidly obese, we might be able to come up with a rational explanation; a bias might not be at work. But as President Biden likes to say: 'Come on, man!' (These words apply as well, I am afraid, to Sugden's suggestion 'that the concept of error-correction, as used in behavioural welfare economics, is not philosophically coherent'.)

Sugden likes the idea of ecological rationality, and he believes that many heuristics are fast and frugal: 'simple rules by which human beings can make the best use of their limited cognitive powers in navigating complex environments'. We agree! I like the idea of ecological rationality too, and I believe that many heuristics are fast and frugal. This claim lies at the heart of the research program inaugurated by Amos Tversky and Daniel Kahneman. Even so, heuristics that generally work well can get us into a heap of trouble (Kahneman *et al.*, 2021); in some cases, they can cost us our lives. A great deal can be done to help (Hu, 2021), and much of the help can be introduced in a Hayekian spirit.

Speaking of that spirit: Regulators all over the world are giving careful consideration to fuel economy and energy efficiency standards (Sunstein, 2021). Such standards promise to reduce externalities, including greenhouse gas emissions. At the same time, aggressive standards are quite expensive, and in cost–benefit terms, they might be challenging to justify solely by reference to externalities; it might be necessary to invoke consumer savings as well. But on standard economic grounds, it is not at all clear that those savings should count. After all, consumers can buy fuel-efficient vehicles if they wish, and if they don't, they must not wish. Why force them? The behavioral response is based an empirical hunch: Consumers might be making errors, perhaps because of present bias, perhaps because of limited attention. A great deal of empirical work investigates that hunch (Sunstein, 2021). The jury is still out, but in my view, the hunch has significant support, in the sense that under epistemically favorable conditions, consumers would be making more fuel-efficient choices (Sunstein, 2021).

https://doi.org/10.1017/bpp.2021.19 Published online by Cambridge University Press

Sugden dips his toe, very lightly, into the empirical waters, investigating one of the many studies. Gillingham *et al.* (2019) study the effect of an abrupt change in the fuel economy ratings of certain vehicle models, brought about by a correction of erroneous ratings. The question was whether the change would produce different choices. The authors find that consumers seem to value $1 in discounted fuel savings at some fraction of that, somewhere between $0.15 and $0.38 – which suggests that some kind of bounded rationality, perhaps present bias, is accounting for their choices. Sugden responds:

> But when an ordinary human consumer is thinking about the future cost of driving a particular model of car, her knowledge of the official ratings will be only one of many disparate ideas that come to mind. She consumer may recall her own experience of the fuel consumption of cars that are in some respect similar to the new model, or what other people have told her about their driving experiences, or what she has read in newspapers and magazines. She may think that her style of driving is such as to make her fuel consumption greater or less than the average. She may – not unreasonably, given what happened in 2012 – be skeptical about the accuracy of official ratings.

This sounds like an explainawaytion. It's a clever one. Is it right? As an objection to the careful work by Gillingham *et al.*, it's pretty seat of the pants. We might want to test it. (Tacit knowledge exists, and it is important, but beware of modern-day phlogiston.)

Rizzo and Whitman (2020) want to have a fight. They urge that I intend 'to persuade the reader that behavioral economic policy has reached the stage where it either can or plausibly could overcome the problems of inadequate knowledge that the two of us have claimed it faces'. They lament that my article 'only mentions us once in a footnote'. They conclude that I aim 'to persuade the reader that knowledge problems pose no great barrier to behavioral policymaking. However, he has not yet addressed the knowledge problems we have presented in the past.'

Carly Simon, a student of human behavior, famously sang, 'You're so vain/You probably think this song is about you.' (A lot of men thought that Simon's song was about them!) My article is not about Rizzo and Whitman. (For the record, I like their book.) With respect to fuel economy standards, the evidence is not clear, which is why I was cautious. If I asked 'the reader to imagine that the knowledge problems 'have *somehow* been overcome', it was because I was sketching what a defense of fuel economy standards, rooted in Hayekian behavioral economics, might look like, not asserting that those problems have, in fact, been overcome. (If you want a full-throated behavioral defense of such standards, invoking the idea of internalities, you would have to look elsewhere.) To the charge that my essay does not address the knowledge problems that Rizzo and Whitman have presented in the past, I plead nolo contendere. This was not the essay's purpose. (Rizzo and Whitman do not address the arguments about Star Wars that I have presented in the past, Sunstein (2016).)

To be sure, one of my main goals was to show that a Hayekian approach would examine not what planners would do, but what informed, unbiased choosers would do. Rizzo and Whitman are skeptical about that whole enterprise. They insist

https://doi.org/10.1017/bpp.2021.19 Published online by Cambridge University Press

that there is no way of knowing what informed, unbiased choosers would do, and that in light of heterogeneity and the existence of local knowledge, the effort to find out is a kind of fool's errand. But (a request, a hope, a plea) might we avoid that level of dogmatism? Or that level of antecedent, uncomfortably-close-to-theological commitment to conclusions that precede engagement with actual problems?

Some cases are really easy; human beings are uninformed or biased, and choosers do err (Bhargava *et al.*, 2017). GPS devices can help; they respond to a lack of information. Reminders can help; they respond to limited attention. Automatic enrollment can help; it can overcome inertia (Thaler and Sunstein, 2021). A lot of people, none of them fools, have been examining the underlying questions (Bhargava *et al.*, 2017; Goldin, 2017; Allcott *et al.*, 2019; Bernheim & Gastell, 2021). It is uncharitable, and not a good idea, to dismiss their work with adjectives and nouns. Again: Tacit knowledge is definitely important, and Rizzo and Whitman are right to draw attention to it. But beware of explainawaytions, and let's post a sign: *No Phlogiston Here*. (Come on, man!)

Sugden urges that in speaking of Hayekian behavioral economics, I have a secret aim, which is 'to head off the criticism that behavioural welfare economics is unacceptably paternalistic'. Actually, I was motivated by something mundane: curiosity. Hayek is a hero of mine (though I have substantial disagreements with him). It seems to me interesting, and perhaps more than that, to think about how Hayek, or a Hayekian, might respond to behavioral findings about individual error. In my view, the resulting approach – Hayekian behavioral economics – deserves respect. It is worthy of serious consideration, even if some of us, including the present author, would not entirely embrace it (Sunstein, 2020b).

References

Allcott, H. and C. Knittel (2019), 'Are consumers poorly informed about fuel economy? Evidence from two experiments', *American Economic Journal: Economic Policy*, **11**(1): 1–37.

Allcott, H. and C. R. Sunstein (2015), 'Regulating internalities', *Journal of Policy Analysis and Management*, **34**(3): 698–705.

Allcott, H. and D. Taubinsky (2015), 'Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market', *American Economic Review*, **105**(8): 2501–2538.

Allcott, H., B. B. Lockwood and D. Taubinsky (2019), 'Should we tax sugar-sweetened beverages? An overview of theory and evidence', *Journal of Economic Perspectives*, **33**(3): 202–227.

Bernheim, B. D. (2016), 'The good, the bad, and the ugly: A unified approach to behavioral welfare economics', *Journal of Benefit-Cost Analysis*, **7**(1): 12–68.

Bernheim, B. D. and J. M. Gastell (2021), Optimal Default Options: The Case for Opt-Out Minimization. Retrieved from: https://www.nber.org/papers/w28254

Bernheim, B. D. and A. Rangel (2007), 'Toward choice-theoretic foundations for behavioral welfare economics', *American Economic Review*, **97**(2): 464–470.

Bernheim, B. D. and A. Rangel (2009), 'Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics', *Quarterly Journal of Economics*, **124**(1): 51–104.

Bernheim, B. D. and D. Taubinsky (2018), 'Behavioral public economics', in B. D. Bernheim, S. DellaVigna, and D. Laibson (eds), *Handbook of behavioral economics: Foundations*. Amsterdam: Elsevier.

Bhargava, S., G. Loewenstein and S. Benartzi (2017), 'The costs of poor health (plan choices) and prescriptions for reform', *Behavioral Science and Policy*, **3**(1): 1–12.

Conly, S. (2013), *Against autonomy: Justifying coercive paternalism*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/bpp.2021.19 Published online by Cambridge University Press

Dworkin, R. (1985), *Law's empire*. Cambridge: Harvard University Press.
Gillingham, K., S. Houde and A. Benthem (2019), Consumer myopia in vehicle purchases: Evidence from a natural experiment. Working Paper 19/321. Center of Economic Research at ETH [Swiss Federal Institute of Technology] Zurich. Retrieved from: https://ssrn.com/abstract=3389832
Goldin, J. (2015), 'Which way to nudge? Uncovering preferences in the behavioral age', *Yale Law Journal*, **125**(1): 226–270.
Goldin, J. (2017), 'Libertarian quasi-paternalism', *Missouri Law Review*, **82**(3): 669–682.
Hu, Z. 2021), Salience and Households' Flood Insurance Decisions. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3759016&dgcid=ejournal_htmle mail_behavioral:experimental:finance:ejournal_abstractlink
Kahneman, D., O. Sibony and C. Sunstein (2021), *Noise*. New York: Little Brown Spark.
Rizzo, M. and G. Whitman (2020), *Escaping Paternalism*. Cambridge: Cambridge University Press.
Sugden, R. (2021), 'How Hayekian is Sunstein's behavioural economics?' *Behavioral Public Policy*, doi:10.1017/bpp.2021.11, 1–10.
Sunstein, C.R. (2016), *The World According to Star Wars*. New York: Dey Street Books.
Sunstein, C. (2020a), 'Behavioral welfare economics', *Journal of Benefit-Cost Analysis*, **11**(2): 196–220.
Sunstein, C. (2020b), *Behavioral economics and public policy*. Cambridge: Cambridge University Press.
Sunstein, C. (2021), 'Hayekian behavioral economics', *Behavioral Public Policy*, doi:10.1017/bpp.2021.3, 1–19.
Thaler, R. and C. Sunstein (2021), *Nudge: The final edition*. New York: Penguin.

Cite this article: Sunstein CR (2023). 'Come on, man!' On errors, choice, and Hayekian behavioral economics. *Behavioural Public Policy* **7**, 212–218. https://doi.org/10.1017/bpp.2021.19