CAMBRIDGE
UNIVERSITY PRESS

RESPONSE

The unsolved Hayekian knowledge problem in behavioral economics

Mario J. Rizzo[1]* and Glen Whitman[1,2]

[1]Department of Economics, New York University, New York, NY, USA and [2]Department of Economics, California State University, Northridge, CA, USA
*Correspondence to: E-mail: mario.rizzo@nyu.edu

> '*No one sees farther into a generalization than his own knowledge of the details extends.*'
>
> William James (Barzun, 1983, p. 14).

There is a limit to the productive exchange of generalizations about public policy. At some point, as William James reminds us, we must go beyond an initial insight or generalization and get into the weeds. This is what we plan to do in our response to Cass Sunstein's article 'Hayekian Behavioral Economics'.

While an exegesis of just where Hayek himself would draw the limits of permissible government intervention may be interesting, this is not the main point of Sunstein's article. What he intends is 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 (Rizzo & Whitman, 2009a, 2020). While Sunstein mentions us only once in a footnote, we are not aware of other attempts to elaborate this 'Hayekian knowledge problem' in detail.[1] We do not point this out because we are desirous of citations, but because it is indicative of a failure to seriously and comprehensively address the relevant issues. In our prior work, we have laid out a series of *specific* knowledge problems that behavioral policymaking – particularly of the paternalist stripe – must confront. Sunstein's latest work does little to address them.

What is the Hayekian knowledge problem?

The original inspiration for Hayek's arguments about knowledge came from the debate in the 1930s and 1940s about the feasibility of rational central planning. The Hayekian argument centered on the question: *How can an economic system*

---

[1]One notable exception is Devereaux (2019).

mobilize knowledge that exists in dispersed bits throughout society such that individuals can use such knowledge to coordinate with each other? His answer is ultimately a comparative institutions claim: the market does a far better job than central planners. Thus, part of the argument is about the merits of the market, and part of the argument is about the knowledge limitations of would-be planners and expert economists. Thus to argue that the market is deficient relative to some abstract ideal – like the perfectly competitive economy – entirely misses his point. This is what Harold Demsetz called the 'Nirvana fallacy' (Demsetz, 1969). To argue that *if* an idealized central planner had all of the requisite information for planning *then* he would perform better than the market is to *assume away* the central problem. The real question is, *how might a planner come to possess such knowledge in the first place?*

Sunstein's defense of behavioral policymaking makes a very similar error. His argument largely relies on a series of contingent claims and speculative situations: *if* we knew all these things, *then* why wouldn't we base policy upon them?[2] Repeatedly, Sunstein mentions potential objections based on the difficulty of obtaining relevant knowledge – only to wave those objections away, asking the reader to imagine they have *somehow* been overcome so his argument may proceed.[3] This approach assumes away the central problem.

To get a better grasp of the knowledge problem for our purposes, we must understand that it is rooted in a distinction between general scientific knowledge or theories and the concrete knowledge of the circumstances of time and place.

The economist, for example, may know or theorize that a firm minimizes its costs to maximize its profits. In many models, the demand curve, the cost of inputs, and the production function are 'given' as data, and the economist merely applies the simple mathematics of constrained optimization. However, the crucial question is: Where does all this data come from, and how would a planner get access to it? The owners and managers of firms come to hold such information *because they have both the means and the motive to acquire it*. Means, because they have direct access to factors of production and direct contact with relevant circumstances. Motive, because they will directly benefit or suffer from the resulting choices.

Hayek further realized that, outside the rarefied conditions of perfect competition, firms differ in their environments. The cost structure faced by one firm may not be the same as that faced by others. In a world of differentiated products, firms will face

---

[2]For example: '*It might well turn out to be possible* to identify those choices [of informed and unbiased choosers]. *Once we have done so*, we might be on the road toward identifying appropriate interventions, whether they involve nudges, taxes, subsidies, or mandates' (Sunstein, 2021, p. 16, emphasis added). And another: 'Such questions raise a host of empirical issues. But *if consumers are not paying enough attention* to savings in terms of money and time, a suitably designed fuel economy mandate might well be justified… *If the benefits of mandates greatly exceed their costs, and if there is no significant consumer welfare loss* (in the form, for example, of reductions in safety, performance, or aesthetics), then the mandates would seem to serve to correct a behavioral market failure' (Sunstein, 2021, pp. 13–14, emphasis added).

[3]For example: 'Notwithstanding Hayek and Mill, some kind of planning institution, or choice architect, might have more information than choosers do, and might not fall prey to a behavioral bias. It is important, of course, to emphasize the word 'might'. Planners might claim to find a behavioral bias when there is no such thing; planners might have (bad) incentives of their own; planners might be subject to the influence of well-organized private groups; planners might themselves be subject to behavioral biases. Still, a corrective response would seem to be on the table' (Sunstein, 2021, p. 7).

different demand curves even in the same industry. And so on. Each firm has concrete knowledge of its time and place. Of course, often firms are not certain about these circumstances, and sometimes they may be wrong. But practically speaking, the central planner will not know any better and likely much worse. His incentives are weaker, and his access to knowledge more indirect and attenuated. A central planner might have knowledge of general tendencies; he might know, for instance, the average explicit cost of production in some industry. But how is that knowledge relevant to the particular plant owner, whose cost structure may differ substantially from the average firm? It is the particular facts of the plant owner's cost structure – what Hayek called *local knowledge* – that matter most.

Aside from being local, relevant knowledge is often *tacit*; i.e., implicit and not easily conveyed to others. Individuals can have a vague sense of their own capabilities. They may know how to do something without being able to explain how to others. For instance, an owner or manager may have a loose sense of how much they could ramp up production in a short period of time, and how much doing so would affect workers' morale. They might not be able to convey that knowledge to anyone else except in the vaguest of terms. Yet such knowledge is no less relevant to the decisions they must make.

The same general principles – comparative analysis, local knowledge, and tacit knowledge – apply to the claim that behavioral planners can create policies that induce individuals to better satisfy their true preferences. However, as we shall see subsequently, the difficulties in this case are even more profound.

A typology of knowledge problems in behavioral policy

In this section we shall assume, for argument's sake, that behavioral paternalists have general scientific knowledge that may be unavailable to individuals. Thus, they know that in experimental studies, phenomena such as present bias, framing susceptibility, base-rate neglect, and so forth have been found. However, the instantiation of these biases is heavily context-dependent (Rizzo & Whitman, 2020, pp. 192–8; 220–3). Even where a bias is instantiated, its effects are also highly contextual (Rizzo & Whitman, 2020, pp. 253–65). These contexts are usually multidimensional and not simply described. This is a manifestation of a more general property of empirical generalizations in psychology: '… few psychological concepts intended to represent a person's tendency to react in a certain way apply across diverse settings' (Kagan, 2012, p. 4). The norm is for cognitive and behavioral tendencies to have relatively narrow domains of application.

The following typology provides an overview of the concrete facts that must be discovered for a behavioral policymaker to have a reasonable chance of improving on the decisions that individuals would make for themselves. In presenting this typology, we are recapitulating lists in our earlier work (Rizzo & Whitman, 2009a, 2020).

Knowledge of true preferences

The fundamental thing policymakers need to know is what people would choose for themselves under idealized conditions wherein they had no relevant information deficiencies, computational limitations, or lack of willpower. This is the Nirvana against

which behavior is to be judged. Practically speaking, however, policy can *at best* move people closer to their optimum. Can this optimum be identified? What evidence is there that people always have unambiguous preferences even under idealized conditions? We know of none. Even if they do exist, can the economist extract these purified or idealized preferences out of a mass of complex data?

Much of the evidence supporting the existence of biased or irrational behavior comes from *inconsistent* behavior. That is, people make different decisions in (what the analyst deems to be) equivalent choice situations. It is not obvious that such inconsistencies are necessarily irrational (Rizzo, 2019). The supposedly equivalent situations may be regarded as different by the individual for legitimate reasons, or she may simply not have settled her mind yet (Whitman and Rizzo (2015), p. 418–19). But even if inconsistencies indeed indicate irrationality, an inconsistency can typically be resolved in more than one way. For instance, if a person exhibits more than one rate of time discount, a more patient one and a less patient one, this fact does not tell us which rate of time discount is the *correct* one (Whitman & Rizzo, 2015, p. 409). The inconsistency could be 'fixed' by inducing the individual to act consistently on the basis of the more patient rate *or* the less patient one *or* some in-between rate. Thus, even if the behavioral planner knows there is an irrationality, he still may not know the fully rational choice.

Knowledge of the extent of bias

It is not enough to know that a bias – such as, for instance, present bias – exists. We need to know its extent in order to calibrate policies like the optimal sin tax. A larger bias presumably requires a stronger correction than a small one. A too-large corrective policy can overshoot the target, creating losses in the opposite direction. Thus, policymaking requires a relatively precise measurement of the size of bias. Unfortunately, economists have had not had much success in measuring discount rates. They tend to run the gamut (Frederick *et al.*, 2002; Cohen *et al.*, 2020). The extent of bias matters for other behavioral policies as well, unless we assume the existence of policies that can simply turn off a bias like a switch.

Knowledge of self-debiasing and small-group debiasing

People are often aware of their biases at least to some degree. They engage in such behaviors as making personal resolutions, avoiding places where they may be exposed to sin goods, adopting systems of self-reward and self-punishment, enlisting the support of family and friends, and choosing diets with allowable exceptions that make the diet palatable.

Self-regulatory schemes are idiosyncratic. The methods that work for one person often will not work for another. In addition, self-regulatory schemes often rely on environmental cues and triggers. In other words, self-regulation involves a great deal of *local knowledge*, just as production decisions do.

Sometimes, these self-regulating behaviors are hard to observe – or can even be mistaken for lack of self-regulation. If a person decides to eat desserts on weekends as part of her diet, we may think that person is inconsistently breaking her plan when, in fact, she is following it. In other words, self-regulation often involves *tacit knowledge* as well.

Yet a paternalist *must* take self-regulatory behaviors into account when constructing policies to improve behavior. Why? Because the presence of self-regulation means the *operative* level of bias may be quite different from the level of bias observed in laboratory studies or even field experiments conducted under different conditions. Without any knowledge of self-regulation, the behavioral planner cannot even conclude that real behavior does not represent true preferences or a close approximation thereof.

Oftentimes people lack knowledge of how to make difficult decisions such as saving for retirement. And yet individuals are not atomistic decision-makers. They discuss things with family and friends. They can get suggestions from trusted sources. There is considerable evidence that when decisions are made in small groups, biases tend to be eroded (Charness & Sutter, 2012; Rizzo & Whitman, 2020, pp. 213–18). So, are the observed decisions arrived at after discussion or consultation? As with self-regulation, this matters for measuring the extent of operative bias in real-world decisions.

Knowledge of counteracting behaviors

As Adam Smith told us, people are not like pieces on a chessboard. They have their own principles of movement or motivation (Smith, 1976 [1759], pp. 233–4). When paternalist policies are narrowly focused to change specific behavior target, individuals often compensate by changing their other actions.[4] For example, there is no reliable evidence that the reduction in sugar-sweetened soft-drink consumption results in weight loss or the reduction of metabolic disorders (Peters & Beck, 2016). One reasonable hypothesis is that noncaloric soft drinks – the likely substitute – do not satisfy the hunger sensation and so people compensate by consuming more calories elsewhere (Markey *et al.*, 2016). Or they may simply treat the reduction in sugary sodas as having 'bought' them some indulgences. So, the paternalistic policy of taxing sugary soft drinks must take account of counteracting behaviors if it is to attain its ultimate goal.

Knowledge of the dynamic impacts on self-regulation

External regulation and self-regulation are substitutes (Fishbach & Trope, 2005). As a result, individuals may respond to greater external regulation, such as sin taxes or nudges, by reducing their level of self-regulation. Aside from reducing the immediate efficacy of the policy, the reduction in self-regulation can have dynamic effects. When internal control is not exercised because of the substitution of prohibitions (or nudges or sin taxes or other policies), the fund of internal self-control capacity deteriorates in the longer run. In one study (Gailliot *et al.*, 2007), those who were asked to practice self-control-related activities for two weeks had better self-control even in an unrelated area afterward. Interestingly, those with the lowest self-control propensities were those who gained most from the exercise of self-control. Thus, the paternalist

---

[4]In other contexts, the tendency to focus on a narrowly defined decision is considered the behavioral bias of 'narrow bracketing'. This is generally defined as the tendency to make decisions without full consideration of other choices and constraints.

needs to see the further implications of his or her attempt to provide a substitute for individuals' own self-control, as it may contribute to the conditions that seemingly call for greater regulation.

Knowledge of bias interactions

It is well known that behavioral psychologists and economists have usually studied the effects of only one bias at a time (Fang & Silverman, 2006). More recently, it has been found that biases are often correlated with each other (Stango & Zinman, 2020). While the implications of bias correlation for behavior have not been adequately explored, we can get some sense of the complexities involved through a plausible example. Present bias tends to reduce the amount of effort a person applies when the burdens are upfront and the rewards down the road. But if the same person is overconfident in the efficacy of his or her efforts and, thus, overestimates the future rewards, it can induce greater effort. We therefore have a case of offsetting biases. The resulting behavior will likely be closer to the rational 'target' than it would with either bias alone. Moreover, the correction of one bias – while leaving the other untouched – can actually make the problem worse. This is an application of the 'second-best' principle to the intrapersonal context (Besharov, 2004).

The possibility of interacting biases has mostly been ignored by would-be behavioral policymakers – *except* when it suits their purposes. For instance, graphic images and risk narratives about the negative consequences of smoking have been used in public service campaigns to offset the deficiency of willpower that smokers tend to exhibit. These images and narratives are exaggerations of the truth that use availability bias to counteract a lack of self-control (Jolls & Sunstein, 2006). Thus, the policy correction works (if it does) by exploiting one bias to correct another.

And yet, even this is an extreme simplification of the phenomenon. In the wild, multiple biases can interact in ways that complicate policy. Smokers already believe that cigarette smoking is more dangerous than it really is (Viscusi, 1990).[5] However, they are alleged to suffer from an irrational optimism bias – the individual smoker believes that he is specially protected from the overall risk. The policy recommended is to frighten a person with exaggerated stories and gruesome images that suggest (through availability bias) higher probabilities of harm than are actually the case. So now there are three alleged biases in play. What must the policymaker know to craft optimal policy under these circumstances? He must know the extent of all three biases as well as how they interact (Rizzo, 2016).

Knowledge of population heterogeneity

For all of the above categories of knowledge, there is substantial heterogeneity. People differ greatly in their 'true' preferences, their exhibited biases, their extent of bias, their self-regulatory strategies, and so on. This is where the rubber truly meets the road, so to speak, with respect to local and tacit knowledge. When people differ in all these ways, the behavioral policymaker's scientific knowledge of biases at the population level is akin to the aforementioned central economic planner's knowledge

---

[5]For the view that subjective beliefs about the health risks of smoking are equal to the objective risks, see Khwaja *et al.* (2007).

of the average cost of production in an industry. An abstract knowledge of the existence of some bias cannot substitute for a specific knowledge of whether and to what extent that bias is instantiated for particular individuals in particular circumstances.

Recent research demonstrates that heterogeneity in the number of biases exhibited by individuals is substantial (Stango & Zinman, 2020). Even if we focus entirely on one type of bias, we cannot assume everyone is biased in the same direction. For example, we cannot assume that a person with an intertemporal bias necessarily has a *present* bias. In a recent NBER study (Stango & Zinman, 2020, pp. 45–6), 29% of the sample was present-biased and 37% was future-biased with respect to money discounting. In the context of food discounting (i.e., healthful food now or later), 15% exhibited present bias, while 7% showed future bias. Even these figures do not give a sense of the heterogeneity within each group (e.g., variation in the extent of present bias among those who have it).

Heterogeneity matters for policy. In an ideal world, the omniscient behavioral planner would create a specific corrective policy for every individual. In reality, the planner must design policies in 'one-size-fits-all' terms. Within a given statutory scheme, everyone faces the same sin taxes, the same nudges, the same attempts to redirect their attention – regardless of their individual level of bias(es). The result will necessarily be a complex combination of undercorrections and overcorrections. To determine whether that is better than inaction, the planner would need a comprehensive knowledge of the entire distribution of biases, bias extents, and so forth – over the whole affected population – as well as a means of weighing the gains for some against the losses for others.

Furthermore, when individuals are categorized into demographic groups, the variation within these groups is even greater than across the groups (Stango & Zinman, 2020, pp. 11–12). This suggests that there is no easy way to avoid the bias discovery problem by substituting well-known demographic categories like income, education, etc., as proxies for biases.[6] Attempts to cater policies to specific groups – assuming such policies could pass muster on ethical–legal grounds of equal treatment – would founder on the same problem of intra-group heterogeneity. The one-size-fits-all-problem for policy is unavoidable. This contrasts with self-debiasing and small-group debiasing, which are in an important sense 'bespoke': they are adopted by private individuals to respond to their own self-identified problems.

Efforts to cope with knowledge problems

There have been previous attempts to address the knowledge problems of behavioral policymaking. In the article under discussion, Sunstein follows the approach of Beshears *et al.* (2008) and Goldin (2015)[7] to construct a series of questions that he believes go some distance toward solving the knowledge problem. Logically, there are two issues: First, are the questions relevant? Second, are the answers known?

---

[6]This creates problems for Goldin's 'Demographic Extrapolation Approach' to uncovering true preferences (Goldin, 2015). See Rizzo and Whitman (2020, p. 274, n. 25).

[7]For a critique of Goldin, see Rizzo and Whitman (2020, pp. 271–74).

What do consistent choosers, unaffected by clearly irrelevant factors or 'frames', choose?

The term 'consistent' choosers is biased or conclusory (in part based on the ambiguous dictionary definition of the word). What is being referred to is *constant* or invariant choosers. Choosers who do not meet this criterion are not necessarily inconsistent with respect to the expression of their true preferences. That is exactly the issue to be decided. We should reiterate that having preferences that are inconsistent (according to the analyst) need not imply irrationality.

The next challenge is identifying 'clearly irrelevant factors'. We suspect that Sunstein would label many more factors as irrelevant than we or other economists would. In the case of the retirement savings defaults, for example, we do not believe that the default, even in the absence of transaction costs, is irrelevant. Defaults can be, and are, interpreted as recommendations from trusted sources (McKenzie *et al.*, 2006). Individuals who may find the decision process complex or otherwise burdensome have their decision-making costs reduced by defaults. Thus, even standard, fully rational neoclassical actors would be affected by these clearly *relevant* factors.

In an important sense, the framing of this question contains a tautology. Decisions unaffected by irrelevant factors are, by definition, more informative. A better question is: How do we know which factors are genuinely irrelevant? And what if the answer is not the same for everyone? Again: heterogeneity matters.

What do informed choosers choose?

Informed about what? To know what information is relevant and whether particular individuals have it requires at least some knowledge of their true preferences. So the analyst is caught in a bit of a circle here. Furthermore, individuals do not differ solely by the degree of knowledge that they possess. For example, it is well established that people who have greater education and higher current incomes are more likely to be financially literate (Tang & Lachance, 2012). And, therefore, they are more likely, based on the two former characteristics *alone*, to prefer enrollment in a retirement-savings program. However, this does not imply that people in the lowest income quartile or a large percentage of them, even when well-informed, would prefer to participate. Their current income is low; they may not have the same preferences for sacrificing current consumption as those with higher current income. Therefore, the preferences of well-informed choosers do not necessarily track the preferences of ill-informed choosers. Privileging the choices of informed choosers does not take the problem of heterogeneity seriously.

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.)

This would be a good question were it not for the 'embarrassment of riches' that characterizes the behavioral research agenda. There are very many biases, aside from framing, and behavioral research suggests that people have them in different environments. Suppose we require people to make an active choice about retirement options. Do they make these decisions free of biases? Surely, out of the more than 175 biases recorded in Wikipedia's page of cognitive biases (Wikipedia, 2021), these people are

likely to have some mix of them with an indeterminate pull in one direction or another. Are their choices purified in the sense required by behavioral economics? How would we even know? The broader point is that active choosing eliminates, at best, the impact of inertia and procrastination. Other biases could prove even more resistant to being eliminating via experiment; eliminating *all* operative biases would pose an even more forbidding challenge. In the world of behavioral economics, a bias-free state of affairs is improbable.

In circumstances in which people are free of (say) present bias or unrealistic optimism, what do they choose?

As with the previous question, the fly in the ointment here is bias interaction. Even the phrasing of the question – that telltale 'say' – suggests the problem. Suppose individuals have both present bias and unrealistic optimism; then by the considerations mentioned earlier, these two biases may counteract each other (by the mechanism described earlier). In general, if people have multiple biases, then any approach that tries to discern 'true' preferences by fixing just one of those biases will be insufficient – and might even produce less desirable choices. The behavioral policy goal effectively requires the simultaneous elimination of all biases to find its target.

What do people choose when their viewscreen is broad, and they do not suffer from limited attention?

The second half of this question cannot be taken literally. We all 'suffer' from limited attention. It is a scarce resource. The only issue is whether our attention is allocated in a way that does not do us harm, all things considered. This cannot be determined by focusing on only one decision or a set of options. People have many things on their minds – perhaps they are focused on their health or children's education. The implication of this is that we can be pretty sure that in real-world (that is, nonlaboratory) cases, individuals will rarely if ever be devoting their *complete* attention – whatever that really means – to any single decision. So we are *always* dealing with limited attention. When is the degree of attention devoted to a decision deemed adequate in behavioral terms?

Attention is a slippery concept and difficult to measure outside of a carefully structured laboratory context. In combination with the emphasis on information in the second question above, ill-measured attention can generate just about any answer to this last question that is desired. If an individual is considered well-informed but decides in a way contrary to the analyst's expectations, the claim can always be made that the individual was not attentive enough to the information she in some sense had.[8]

Behavioral intervention in principle and practice

When Sunstein and Richard Thaler first introduced behavioral paternalism in the early 2000s, they expressly advocated 'softer', less intrusive, and less coercive means

---

[8]For a discussion of the related concept of salience, see Rizzo and Whitman (2020, pp. 158–62) and for a discussion of the possible confusion of salience with novelty, see Id. at p. 258.

than old-style paternalism (Sunstein & Thaler, 2003; Thaler & Sunstein, 2008). Although they said soft and hard paternalism lay on the same spectrum and there was no sharp line to separate them (Sunstein & Thaler, 2003, p. 1185), they nevertheless stressed the virtue of less intrusive measures, including new default rules and waivable terms, that would ostensibly preserve freedom of choice. The terms 'libertarian paternalism' and 'nudge' were both chosen to make this point.

Yet, in the present article, Sunstein's leading example is fuel economy mandates that cannot be avoided or waived in any fashion. He explicitly defends mandates and bans as superior to less coercive measures: 'But we could certainly identify cases in which the best approach is a mandate or a ban, because that response is preferable, from the standpoint of social welfare, to any alternative, including information, economic incentives, or defaults' (p. 10). And he impugns freedom of choice head-on: 'If we know that people's choices lead them in the wrong direction, why should we insist on freedom of choice?' (p. 6).

In earlier work (Rizzo & Whitman, 2009b), we argued that Sunstein and Thaler's theoretical framework had an inherently expansive tendency, and that early signs of movement in that direction were already apparent. The present article seems to confirm our prediction.

Whether the policies in play are 'soft' or 'hard' matters for the knowledge problem. Why? Because more intrusive and coercive policies presumably bear a larger burden of proof. The planner has to know more about all relevant valuations to conclude, with confidence, that some options should be foreclosed entirely.

The new behavioral paternalism also purported to be scientifically grounded and evidence-based. Sunstein and Thaler (2003, p. 1166) say that interventions 'should be designed using a type of welfare analysis, one in which a serious attempt is made to measure the costs and benefits of outcomes …' (Sunstein & Thaler, 2003, p. 1166). In the article under discussion, Sunstein (2021, p. 14) says that 'Behavioral biases have to be demonstrated, not simply asserted …' All of this is in contrast to the old paternalism that seemed to care little for such scientific discipline.

On the other hand, there is a parallel line of argument evident in previous work (Sunstein, 2015) that relies on the premise that departures from rationality are sometimes (often? usually?) obvious, as shown by seemingly simple examples. We have dubbed this approach 'the appeal to obviousness' (Rizzo & Whitman, 2020, p. 407). The examples used gain credence through a construction that assumes away the knowledge problems we have reviewed here. This is clear from the examples of obvious irrationality given by Sunstein (2015, pp. 517–18). Let us consider just two:[9]

1. 'Jones is asked to make a choice between two identical radios. One costs more. He chooses the more expensive one.'
2. 'Jones is buying a car. The first option costs very slightly less than the second but has terrible fuel efficiency, so much so that after 6 months, the second option would save him money. (He expects to own the car for at least 5 years.) He likes the two cars the same. He chooses the first option, because he pays no attention to the fuel economy of the cars.'

---

[9]For a complete analysis of these examples and others, see Rizzo and Whitman (2020, pp. 407–11).

As presented, these examples have the character of a tautology. They are set up so as to be necessarily but trivially irrational. The two radios are identical to Jones, by assumption. The two cars are the same to Jones except for fuel efficiency, by assumption. Jones plans to keep his car for long enough to make the extra upfront cost worth it, by assumption. And so on. But in the real world, it is precisely on questions like these, relating to the true perceptions and valuations of customers, that knowledge problems are most intractable.

To illustrate, let us examine the real-world case of the fuel-economy mandate that Sunstein presents as a hypothesis. When we actually look at the evidence, what do we find? Allcott and Knittel (2019) provided subjects with information about fuel-economy savings for automobiles in what seemed to be highly salient terms. The information was presented in terms of concrete trade-offs (e.g., trips to Hawaii), and subjects had to pass a quiz to make sure they understood, thereby ensuring that sufficient attention had been paid. The conclusion was that there was *no statistically or economically significant effect on purchases.*

The important question is: Why? One obvious answer is that subjects simply valued the other features of less fuel-efficient cars – styling, legroom, acceleration, and so on – more than the fuel-efficiency savings, and perhaps more than they reported themselves when surveyed (as they were in this study).

But the authors consider other possibilities. They note that there may have been other biases in play, such as cognitive costs in *using* information, imperfect memory (when they actually bought the car later), possible narrow view in not knowing the fuel costs of the cars not being considered at the time, and a lack of trust in the information. So, maybe these had an effect. To identify what consumers *truly* want, we would need to control for *all* of these factors – and any factors that might point toward *over*-valuing fuel economy as well.

To their credit, the authors simply conclude 'either that some other market failure or behavioral failure must justify the CAFE standard, *or* that the large net private benefits projected in the CAFE Regulatory Impact Analyses do not exist' (Allcott & Knittel, 2019, p. 34, emphasis added).

Consider what a cost–benefit analysis of fuel-economy mandates would involve if knowledge problems were taken seriously. It would consider all the behavioral factors above, including offsetting biases. It would take into account the extent of these biases to determine how much of the fuel savings should count as a benefit. It would allow for rates of time-discounting that differ from the planner's. It would consider the value that consumers place on numerous other features of cars. And most importantly, it would consider the heterogeneity of all these factors and more. As Gayer and Viscusi (2013) show, actual fuel-economy analyses by the Environmental Protection Agency and National Highway Traffic Safety Administration do not even approach this level of sophistication – and in particular, that they effectively ignore the problem of preference heterogeneity. Such heterogeneity is a problem for all one-size-fits-all policies, but *especially* mandates that allow no exit options.

How does Sunstein handle these concerns? Simply put, he assumes them away. For instance, here is his only reference to heterogeneity: 'As noted, there might be heterogeneity in the relevant population, making it challenging to generalize from what some part of a population does. *But suppose that there is no such heterogeneity.* In

principle and sometimes in practice, efforts to answer these subsidiary questions should help public institutions with welfare analysis …' (Sunstein, 2021, p. 9, emphasis added).

When Sunstein accuses standard economists of 'explain-away-tions' – attempts to explain away behavioral problems by asserting arbitrary *a priori* rationalizations of deficient behavior (Sunstein, 2021, p. 7, n. 3) – he appears to be taking a scientific stance. But when we examine the evidence in this case, we see that the alternate explanations are scientifically plausible and that the behavioral explanations have not been demonstrated. In the face of serious knowledge problems, Sunstein replies to 'explain-away-tions' with 'assume-away-tions'.

Conclusion

Sunstein's latest work aims 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 (and that we recapitulate here). The mechanisms he suggests that behavioral planners can use to acquire the necessary knowledge do not pass muster. At best, they access some specific knowledge about particular types of individuals under particular circumstances, without the necessary breadth and depth needed to generalize the results to a heterogeneous population. Without such comprehensive and centralized knowledge, we cannot predict that proposed interventions will indeed improve upon the decisions that people make for themselves on the basis of their own local and tacit knowledge.

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.

Barzun, J. (1983), *A stroll with William James*. New York: Harper & Row.

Besharov, G. (2004), 'Second-best considerations in correcting cognitive biases', *Southern Economic Journal*, **71**(1): 12–20.

Beshears, J., J. J. Choi, D. Laibson and B. C. Madrian (2008), 'How are preferences revealed?', *Journal of Public Economics*, **92**(8–9): 1787–94.

Charness, G. and M. Sutter (2012), 'Groups make better self-interested decisions', *Journal of Economic Perspectives*, **26**(3): 157–76.

Cohen, J., K. M. Ericson, D. Laibson and J. M. White (2020), 'Measuring time preferences', *Journal of Economic Literature*, **58**(2): 299–347.

Demsetz, H. (1969), 'Information and efficiency: Another viewpoint', *The Journal of Law and Economics*, **12**(1): 1–22.

Devereaux, A. N. (2019), 'The nudge wars: A modern socialist calculation debate', *The Review of Austrian Economics*, **32**(2): 139–58.

Fang, H. and D. Silverman (2006), 'Distinguishing between cognitive biases', in E. J. McCaffery and J. Slemrod (eds), *Behavioral Public Finance*, New York: Russell Sage Foundation, 47–81.

Fishbach, A. and Y. Trope (2005), 'The substitutability of external control and self-control', *Journal of Experimental Social Psychology*, **41**(3): 256–70.

Frederick, S., G. Loewenstein and T. O'donoghue (2002), 'Time discounting and time preference: A critical review', *Journal of Economic Literature*, **40**(2): 351–401.

Gailliot, M. T., E. A. Plant, D. A. Butz and R. F. Baumeister (2007), 'Increasing self-regulatory strength can reduce the depleting effect of suppressing stereotypes', *Personality and Social Psychology Bulletin*, **33**(2): 281–94.

Gayer, T. and W. K. Viscusi (2013), 'Overriding consumer preferences with energy regulations', *Journal of Regulatory Economics*, **43**(3): 248–64.

Goldin, J. (2015), 'Which way to nudge: Uncovering preferences in the behavioral age', *Yale Law Journal*, **125**: 226.

Jolls, C. and C. R. Sunstein (2006), 'Debiasing through law', *The Journal of Legal Studies*, **35**(1): 199–242.

Kagan, J. (2012), *Psychology's ghosts: The crisis in the profession and the way back*. New Haven: Yale University Press.

Khwaja, A., F. Sloan and S. Chung (2007), 'The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions', *Journal of Risk and Uncertainty*, **35**(2): 179–201.

Markey, O., J. Le Jeune and J. A. Lovegrove (2016), 'Energy compensation following consumption of sugar-reduced products: A randomized controlled trial', *European Journal of Nutrition*, **55**(6): 2137–49.

McKenzie, C. R., M. J. Liersch and S. R. Finkelstein (2006), 'Recommendations implicit in policy defaults', *Psychological Science*, **17**(5): 414–20.

Peters, J. C. and J. Beck (2016), 'Low calorie sweetener (LCS) use and energy balance', *Psychology and Behavior*, **164**(B): 524–8.

Rizzo, M. J. (2016), 'The four pillars of behavioral paternalism', in S. Abdukadirov (ed.), *Nudge theory in action: Behavioral design in policy and markets*, New York: Palgrave, 37–63.

Rizzo, M. J. (2019), 'Inconsistency is not pathological: A pragmatic perspective', *Mind & Society*, **18**(1): 77–85.

Rizzo, M. J. and D. G. Whitman (2009a), 'The knowledge problem of new paternalism', *BYU Law Review*, **2009**(4): 905–67.

Rizzo, M. J. and D. G. Whitman (2009b), 'Little brother is watching you: New paternalism on the slippery slopes', *Arizona Law Review*, **51**: 685–739.

Rizzo, M. J. and G. Whitman (2020), *Escaping paternalism: Rationality, behavioral economics, and public policy*. Cambridge, UK: Cambridge University Press.

Smith, A. (1976 [1759]), *The theory of moral sentiments*, D. D. Raphael, and A. L. Macfie (eds). Oxford, UK: Oxford University Press.

Stango, V. and J. Zinman (2020), *We are all behavioral, more or less: A taxonomy of consumer decision making (No. W28138)*. Cambridge, MA: National Bureau of Economic Research.

Sunstein, C. R. (2015), 'Nudges, agency, and abstraction: A reply to critics', *Review of Philosophy and Psychology*, **6**(3): 511–29.

Sunstein, C. R. (2021), 'Hayekian Behavioral Economics', *Behavioural Public Policy*, 1–19. doi:10.1017/bpp.2021.3.

Sunstein, C. R. and R. H. Thaler (2003), 'Libertarian paternalism is not an oxymoron', *The University of Chicago Law Review*, **70**(4): 1159–202.

Tang, N. and M. E. Lachance (2012), 'Financial advice: What about low-income consumers?', *Journal of Personal Finance*, **11**(2): 121.

Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Viscusi, W. K. (1990), 'Do smokers underestimate risks?', *Journal of Political Economy*, **98**(6): 1253–69.

Whitman, D. G. and M. J. Rizzo, (2015), 'Problematic welfare standards of behavioral paternalism', *Review of Philosophy and Psychology*, **6**(3): 409–25.

Wikipedia (2021), List of Cognitive Biases. Retrieved from: https://en.wikipedia.org/wiki/List_of_cognitive_biases (accessed on April 9, 2021).