International Review of Economics (2020) 67:419–441
https://doi.org/10.1007/s12232-020-00349-7

RESEARCH ARTICLE

Libertarian paternalism: taking Econs seriously

D. Wade Hands[1]

Received: 13 August 2019 / Accepted: 25 April 2020 / Published online: 6 May 2020

Abstract

There is an extensive critical literature analyzing the libertarian paternalism (LP) of Richard Thaler and Cass Sunstein. This paper is critical as well, but does so from a different perspective than most of the existing research. Thaler and Sunstein characterize LP by at least two key features: (1) a sharp distinction between Econs (those whose behavior will be unchanged by LP policies) and Humans (who will, at least potentially, change their behavior as a result of LP policies), and (2) defining Econs explicitly as *homo economicus*: "the textbook picture of human beings offered by economists" (Thaler and Sunstein in Nudge: improving decisions about health, wealth and happiness. Penguin, London, p. 7, 2009). This paper will take their definition of Econs seriously and examine the implications for LP-based policies. The bottom line is that if we take Econs seriously, LP nudges end up being not only extremely weak policy tools, but they also fail to accommodate some of the most important insights of behavioral economics.

Keywords Libertarian paternalism · Nudging · Rational choice theory · Behavioral economics · Behavioral welfare economics

JEL Classification D60 · D90

Versions of this paper were presented at the "Norms and Normativity" conference in Lyon, France, June 27–29, 2018, in a symposium at the University of Nice, July 3, 2018, and at the Allied Social Science Meetings in Atlanta, January 4–6, 2019. Helpful comments were received from various members of these audiences and as well as from Magdalena Malecka, Ivan Moscati, and two anonymous reviewers.

✉ D. Wade Hands
hands@pugetsound.edu

[1] Department of Economics, University of Puget Sound, Tacoma, WA 98416, USA

1 Behavioral economics and libertarian paternalism

Some of the ideas associated with behavioral economics have a fairly long history, but one major impetus for the contemporary literature came from the research of Daniel Kahneman and Amos Tversky during the 1970s: their 1974 *Science* paper (Tversky and Kahneman 1974) and their *Econometrica* paper (Kahneman and Tversky 1979) on prospect theory in 1979. Their approach to individual decision making was applied to economic choices by Richard Thaler and other economists giving birth to the *heuristics and biases program* within behavioral economics. The defining feature of heuristics and biases (hereafter HB) research has been to provide empirical evidence that actual human decision makers frequently behave in ways that are inconsistent with rational choice theory; they make *mistakes* and these deviations from rationality are often systematic and repeated.[1] As Thaler recently put it: "The approach taken by most behavioral economists has been to focus on a few important ways in which humans diverge from *homo economicus*" (Thaler 2017, p. 1800). The result has been a vast number of empirical *anomalies*—including loss aversion, framing effects, endowment effects, hyperbolic discounting, anchoring effects, and many others—and while it is certainly possible to criticize some of this research, the sheer number and persistence of these results suggest they cannot be ignored.

Although there was a protracted debate over whether behavioral economics or traditional utility theory is better for predicting and explaining individual choice—and skirmishes still flare up from time to time—for the most part, significant debate over the scientific status of behavioral economics has died down and behavioral economics is now an established part of mainstream economic research and teaching.[2] But this does not mean that controversy has ended. In recent years debate has increasingly turned toward the *normative* implications of behavioral economics. If behavioral economics has demonstrated that economic agents often make mistakes and generate anomalies, it raises serious questions about the relationship between what they *actually* choose and what they *ought to choose* (and there is more than one *ought* to consider). One of the leading topics in these discussions is the literature examined here: *nudging* and in particular the *libertarian paternalism* (hereafter LP) of Richard Thaler and Cass Sunstein (Sunstein 2013, 2015; Sunstein and Thaler 2003; Thaler and Sunstein 2003, 2009) and the related work on *asymmetric paternalism* (Camerer et al. 2003).

LP begins from the HB position that individuals make *mistakes*, cognitive errors, but seeks to find ways to nudge these individuals back to more rational choices *without using either coercion* or *incentive-based economic tools*. Since one of the main messages of the HB literature is that the choice context matters to outcomes, it is

[1] This literature is too extensive to provide comprehensive references, but a few key works include: Camerer and Loewenstein (2004), Kahneman (2003); Kahneman et al. (1991), Kahneman and Tversky (2000) and Thaler (1980, 2000, 2018).

[2] As exhibited by the amount of behavioral economics published in prestigious economics journals and the presence of advanced textbooks such as Dhami (2016).

argued that the individual's choice environment—the choice architecture—can often be changed in ways that will nudge the individual into making better choices. As Thaler and Sunstein explain:

> In our understanding, a policy is "paternalistic" if it tries to influence choices in a way that will make choosers better off, *as judged by themselves*. Drawing on some well-established findings … we show that in many cases, individuals make pretty bad decisions—decisions they would not have made if they had paid full attention and possessed complete information, unlimited cognitive ability, and complete self-control. (Thaler and Sunstein 2009, pp. 5–6)[3]

Two frequently discussed examples are the director of food services for a school system who rearranges the way that cafeteria food is presented so that healthy items are more likely to be selected, and the corporation that changes its opt-in retirement plan to an opt-out system in order to increase plan participation (Thaler and Sunstein 2009). Notice that such changes in the choice architecture are *paternalistic*—they are changes designed to make students and employees better off—but they are also *libertarian* in the sense that people are still free to choose; the less healthy food is still available and employees are still free to opt out of the company's retirement plan.

The changes in the choice architecture are designed so that individuals who are prone to HB-type mistakes will be nudged into more rational choices, while those who are not prone to such mistakes will not change their behavior as a result of LP nudging. Thaler and Sunstein have introduced particular terminology for these two groups; those who do not make such mistakes are called *Econs* (the *homo economicus* of standard economic theory) and those who do make such mistakes are called *Humans* (although *Homo Heuristicus* may have been a better choice). Again, Thaler and Sunstein:

> Whether or not they have ever studied economics, many people seem at least implicitly committed to the idea of *homo economicus*, or economic man—the notion that each of us thinks and chooses unfailingly well, and thus fits within the textbook picture of human beings offered by economists.
> If you look at economics textbooks, you will learn that *homo economicus* can think like Albert Einstein, store as much memory as IBM's Big Blue, and exercise the willpower of Mahatma Gandhi. Really. But the folks that we know are not like that … To keep our Latin usage to a minimum we will hereafter refer to … Econs and Humans." (Thaler and Sunstein 2009, p. 7)

[3] The discussion in this paper will stay very close to Thaler and Sunstein's original definitions, but there exists quite a bit of variation in the way that different authors use the terms nudging and LP. It seems that almost everyone agrees that nudging is the general concept—broadly altering the individual's choice architecture to correct for mistakes in decision making—and that nudging could be done for many different reasons. On the other hand, LP is a special case of nudging concerned with paternalistic goals, but there is disagreement about what exactly makes it special and how it should be defined.

Perhaps Thaler and Sunstein are not serious about *homo* economicus and such remarks are just throw away lines, but it seems reasonable to ask what the implications for LP would be if we took their words seriously and used the textbook *homo economicus*—a clear and well-known model of individual decision making—as the basis for characterizing Econs and Humans, as well as investigating various other aspects of the LP program. When we do this, the exact character of *homo economicus* becomes key to the entire LP program since it carries both descriptive and normative weight. Descriptively it distinguishes the agents who will be, and those who will not be, affected by LP nudges; and normatively it distinguishes the agents whose decision making should be corrected for cognitive errors, and those whose decision making is beyond reproach. If we take what Thaler and Sunstein say about *homo economicus* seriously then it is clear that Econs and Humans are the foundations of the LP program, and it is also clear they are both caricatures: idealized models of preference-based decision making. Like the agents in traditional economics textbooks, Econs are endowed with stable well-ordered preferences that (along with beliefs and constraints) are causally responsible for, and/or can systematically rationalize, the choices of individual Econs. But it also follows that Humans have Econs deep inside—an *inner rational agent* (Infante et al. 2016, p. 14)—but that inner Econ is seldom responsible for, and/or rationalizes, Human choices, because that inner rational agent is surrounded by a *psychological shell* of heuristics, biases, frames, and other factors which systematically prevent Humans from manifesting the preferences of their *inner rational agent*. As Infante, Lecouteux, and Sugden explain:

> … ordinary human psychology is being treated as a set of forces that are liable to restrict the inner agent's ability to act according to the implications of its own reasoning. It is as if the inner rational agent is separated from the world in which it wants to act by a *psychological shell*. The human being's behaviour is determined by interactions between the autonomous reasoning of the inner agent and the psychological properties of the outer shell. However, in relation to issues of preference and judgement, the inner agent is the ultimate normative authority. (2016, p. 14)

In other words, both Econs and Humans have "an ideally rational agent skulking within" (Hausman 2016, p. 26), but for Humans it is an inner agent "whom their actions betray" (ibid.). Thus, even though both Econs and Humans are idealized agents, Econs are foundational since Humans are essentially "faulty Econs" and their normative standard is rational action based on "the preferences of the imagined inner Econ" (Infante et al. 2016, p. 22).

It is useful to note that given these definitions, "Econs or Humans" is an incomplete disjunction; actual living flesh-and-blood humans need not be making decisions as circumscribed by either the Econ or Human model (or for that matter any preference-based model of decision making). As Robert Sugden points out:

> Despite Thaler and Sunstein's label, the decision maker described by this model, is not a Human in the ordinary sense of the word. It is a faulty Econ." (Sugden 2017, p. 117)

The psychological, sociological, and biological literatures of course have many different ways of predicting, explaining, and rationalizing the behavior of individual *homo sapiens* which do not involve preference or utility in any way, and thus are quite different from either Econs or Humans. For example, actual humans may make the choices they do because their behavior is a result of the simple conditioned responses of early behaviorism; or because they are nothing but robotic survival machines being propelled by the replication of their selfish genes (Dawkins 1976); or perhaps behavior is structurally and culturally determined as with the *homo sociologicus* of traditional social theory; or perhaps it is because of the boundedly rational, but not preference-based, mechanisms of fast-and-frugal ecological rationality (Gigerenzer 2008; Gigerenzer and Brighton 2009)[4]; and there are of course many other possibilities.

The point of noting some of the many other ways that scholars have theorized about real human decision making is not to defend non-preference-based ways of explaining human behavior; it is simply to note that LP does not start with observations of actual behavior and identify behavioral regularities that might serve as the guideline for LP policies. Rather, LP is model-driven; it starts with the narrow range of behavior defined by two specific types of idealized choosers: Econs whose behavior is the result of successful satisfaction of well-behaved and stable preferences, and Humans who also have such an inner rational agent, but whose behavior fails to achieve preference satisfaction because of interference from their outer psychological shell. This characterization of LP is admittedly narrow relative to the way that LP is often portrayed in the literature so additional discussion of this conception is provided below.

2 Econs, humans, and nudges: a closer look

It is often noted that advocates of LP are fairly ambiguous about what exactly is, and is not, a LP-based policy (e.g., Hansen 2016; Grüne-Yanoff and Hertwig 2016; Rebonato 2012). To some degree, this is a result of the examples-driven style of much of the LP literature which starts with a few examples of non-incentive-based and non-coercive policies that change behavior in ways that seem obviously good—better health, longer life, more savings, and so forth—and then conduct the rest of the analysis through the lens of these initial examples. Instead of starting with clear definitions and consistent foundational commitments, and ending with policies that reflect those definitions and commitments, the process is the reverse; the discussion begins with certain (presumed to be obvious) exemplars of good policies and proceeds by identifying various concepts that rationalize those exemplars: "building an 'ostensive' rather than 'axiomatic' definition of libertarian paternalism" (Rebonato 2012, p. 6). This has led to a vast amount of LP literature, but also to a certain amount of ambiguity:

[4] See Schmidt (2019) for an interesting effort to build a version of nudging on this version of ecological rationality.

> "Without clear and consistent foundational concepts the new policy paradigm of applied behavioural science may easily come to seem ill founded, leaving the concept of nudge as well as the ideology of libertarian paternalism vulnerable to accusations of slippery-slopes, claims of conceptual inconsistency, and warnings that nudges may quickly turn into shoves …" (Hansen 2016, p. 157)

While this diversity of interpretations may have benefits in terms of wide-ranging policy applications, the examples-driven approach also has costs. For example, it is not exactly clear what constitutes a LP nudge (as opposed to say, a nudge in the social interest, or traditional paternalism), whether the goal is for Humans to make more rational decisions or ones that will make them better off (or perhaps both), and how exactly LP nudges are related to Thaler and Sunstein's commitment to textbook *homo economicus* as the normative baseline for LP nudging. This paper approaches the problem from a different direction—bottom-up rather than top-down—by taking what Thaler and Sunstein say about the role of *homo economicus* seriously and drawing out the implications of using Econ as the normative standard, seeing what constraints it imposes on Humans, and using these foundations to investigate what LP looks like through this lens.

The main reason for optimism about this approach is that, unlike starting with potentially conflicting intuitions about what constitutes good policy outcomes, *Econs are very well-defined* and provide a relatively uncontentious point of departure for the analysis (perhaps not uncontentious with respect to either their scientific or normative adequacy, but with respect to their identity: what they *are* and what kind of agency and normative guidance they support). There has been relative stability within microeconomic textbooks and the associated core commitments of economists about Econ agency since roughly the 1940s; this means *there is a fact of the matter about what it is like to be an Econ*. It has been said that Econs and Humans are a "pleasant but obscure allegory" (Mongin and Cozic 2018, p. 111), but the position taken here is that *what Econs are*—and thus *what Humans must be* because they are faulty Econs—is the *least obscure* aspect of the LP literature. There are of course many other interpretations of what Econs and Humans are like, but the main motivation for this paper is to see what we get when we analyze LP starting with the part of Thaler and Sunstein's characterization that is the least obscure. Given Econs as a well-defined starting point, much of the rest of the paper will play out as a transcendental argument regarding what must be the case for Humans and the associated LP policies so that it is possible for Econ to serve effectively as the normative behavioral standard.

The rest of this section and the following section will discuss a number of features of Econs and Humans that give us a better understanding of the foundations and limitations of LP. The first of these is the difference between *pro-self* and *pro-social* nudging (Barton and Grüne-Yanoff 2015; Hagman et al. 2015). Pro-self-nudges are nudges designed to make agents more effective individual preference satisfiers and more likely to avoid HB mistakes, while pro-social nudges are designed to achieve

social goals.[5] It is often pointed out that while Thaler and Sunstein's definitions make LP exclusively about (private) individual preference satisfaction, the examples they use often cross the line between pro-self and pro-social nudges:

> A significant part of the nudge literature is directed at using behavioural insights to induce "behaviour change" in situations in which the targeted individuals do not seem to be making mistakes in satisfying their own preferences … they are simply frustrating the achievement of some public policy objective. For example, TS's [Thaler and Sunstein] catalogue of emulation-worthy policies includes nudges designed to reduce littering, to increase registration in organ donation programmes and … to reduce the release of potentially hazardous chemical into the environment. (Infante et al. 2016, p. 5)[6]

While the distinction between pro-self and pro-social nudging may often be blurred within the LP literature, the distinction nonetheless provides a very clear analytical way to define LP relative to other types of nudging: *LP is purely pro-self-nudging* (Barton and Grüne-Yanoff 2015, p. 344).

Notice that defining LP in this way (as will be done for the remainder of this paper) introduces yet another way that LP involves idealization since any actual nudge, however pro-self the choice architect intended it to be, is almost certainly going to have some social impact. A purely pro-self-nudge—like a perfectly rational consumer—can be modeled, but it will involve a number of idealizations that will almost never be present in real target applications.

For example, suppose we observe Fred consuming far more unhealthy junk food than seems rational (based on the best medical advice). From this observation it is not obvious whether Fred (1) has well-ordered preferences and is acting optimally on them (is an Econ, or at least approximately one), (2) has well-ordered preferences but is not acting optimally on them for some HB reasons (is a Human, or at least approximately one), or (3) is acting on the basis of other, non-preference-based reason, value, motivation, cause, etc. In a world where Fred could only be an Econ or a Human, the choice architect could implement a LP nudge—assuming they know which HB problem causes such dietary mistakes—and observe whether Fred reduced his consumption of junk food or not. If he did, he would be revealed Human, and if not he would be revealed Econ. The problem is of course that Econ or Human are not the only possibilities when thinking about nudges on real human beings. Given a target population of real humans it very unlikely that the choice architect would be able to predict the outcome of the LP nudge or to understand exactly why those who changed their behavior did so.

One advantage of starting from Econs is that it allows us to clearly see the difference between a pro-self LP nudge and a pro-social nudge. Continuing with Fred,

[5] Of course there may be all kinds of social goals from justice, to freedom, to fairness, and so forth, but for the purpose of this paper social goals mean what they mean in standard microeconomics textbooks: correcting for negative and positive externalities, providing public goods, etc.

[6] Also see Sunstein (2016) where survey questions include policies to "reduce pollution" and "encourage water conservation."

suppose that Fred is fully informed about the health effects of such eating, but puts a very high value on the taste of food and has no particular desire to live a long life. So in this case Fred really does prefer to eat junk food and is acting rationally to satisfy his preferences; he is an Econ, at least with respect to junk food, so there is no room for LP nudging. The textbook characterization of *homo economicus* certainly allows for rational consumption of junk food, or cigarettes, or a variety of other things that most of us would say (with good evidence) are not really good for us, but this is part of what it means to be an Econ. Rationality according to rational choice theory is about rational/optimal satisfaction of one's given preferences and not about what it is rational to prefer.

But even in the case of Econ Fred there might be room for pro-social nudging. We live in a society with a myriad of interdependencies and thus a myriad of possible external effects. There is a high probability that Fred will be unhealthy and require more medical expenditure than the average citizen. That extra cost will be paid in part by other citizens, either through higher insurance costs or higher taxes (or both). There are of course many other possible externalities, but the point is simply that nudge-type policies might well be used to change Fred's junk food consumption for pro-social reasons, but if so, it would not be LP nudging. Of course Thaler and Sunstein say that Econs will not respond to LP nudges, and that is entirely correct; Econ Fred will not respond to changes in choice architecture designed to make him a more effective preference satisfier since he is already behaving rationally. But it may be possible to change the choice architecture in such a way that Fred reduces his consumption of junk food in a way that serves the social interest. This pro-social nudge will of course make him a less efficient preference satisfier—and therefore less rational in the Econ sense—but if the negative externalities are large enough it may be socially beneficial to implement such a policy. Of course, it is also possible that the socially desired change could come from a more traditional policy, like higher junk food taxes, or perhaps even some combination of nudges and more traditional incentives and disincentives.

Finally, this way of thinking about nudges also allows us to analytically differentiate not only between the cases of LP and pro-social nudges, but it also provides some insight into *traditional, or "hard paternalist,"* policies. So now consider Sally. Suppose Sally is fully informed, acts rationally, and really does have a strong preference for junk food—i.e., is an Econ when it comes to junk food—but now suppose she lives alone as a hermit and her eating habits impose no externalities on anyone else in society. In this case *neither* a LP nudge (because she is acting rationally) nor a pro-social nudge (since there are no external costs) is needed, but we still might want to change her eating behavior because eating junk food is—based on our best available evidence—*not good for her*. This is one way to characterize a *traditional paternalist* intervention; the motivation is what is really good for the person and has nothing to do with the individual's preferences or whether they are acting rationally given those preferences. In this case it may be possible to introduce a pure-paternalist nudge, or a more traditional incentive-based policy, that would change Sally's behavior in the direction of what is actually good for her; such an intervention would make Sally really better off, but it would not be LP since it would change her behavior in a way that is contrary to the desires of her inner rational agent.

The bottom line seems to be that we can distinguish at least three different (pure) kinds of nudge-based interventions: LP pro-self-nudges aimed at helping people better satisfy their preferences, pure pro-social nudges based on reducing externalities or the production of public goods, and traditional paternalist nudge which make people better off independently of what they prefer. In reality of course there could be combinations of all three, as well as the possibility of various outcomes and motivations that are not identical to any of these. All this implies that a pure LP nudge will be a very difficult, if not impossible (see section IV below) intervention to even identify, much less execute, if one is consistent with the Thaler and Sunstein definitions of Econs and Humans. This means that LP policies will constitute a *very narrow class of interventions*: a class that is often inconsistent with what those sympathetic to LP say about the range of LP policy applications.

Perhaps an example that is a bit more real world would be useful to help clarify the difference between a pro-self LP policy and the more traditional economic notion of a social policy. To this end, consider a relatively low impact environmental problem like littering. If we think of the problem solely in LP terms, it is only a problem if those littering prefer not to litter and their littering is the result of making various HB-type mistakes that lead them to generate non-utility-maximizing levels of litter. If the problem is viewed strictly in LP terms the role of policy would be to (1) find out whether the individuals in question *really preferred* not to litter, (2) discover what particular heuristic was preventing them from producing the utility-maximizing amount of litter, and then (3) design a nudge that would change the choice environment in such a way that it would lead them to generate less litter. By the way, if, during (1), the preference examination phase, it was discovered that the individuals in question really do *prefer to litter*, then as a pure LP nudger, absolutely nothing should be done to change their behavior. Now consider the problem in a more traditional way. Litter is a negative externality, it imposes external costs on others in the society, and since the cost is not paid by the people who litter, they tend to overproduce it unless there is some disincentive to do otherwise. In this case the policy has a direct reason to reduce litter—it imposes social costs on others in the community—and it is relatively easy to implement since it is fairly easy to detect who is, and who is not, littering. One of the traditional solutions in such a case is simply to put a fine on those who litter and the litter will consequently be reduced, although some sort of social nudge could also be used. It should be noted that when the litter is viewed as an externality rather than a mistake in rational decision making, the litter will be reduced regardless of whether those who littered genuinely preferred to litter or whether they really didn't really want to do it, but couldn't stop themselves because of the interference of their outer psychological shell.

Given that the language of externalities has been introduced it is probably a good time to introduce the language of *internalities*—or internal externalities—a behavioral economics concept originally introduced in Herrnstein et al. (1993). The idea of an internality mirrors the traditional idea of an externality, but it is inside the individual. Focusing for simplicity on the case negative externality, a difference between private and social cost, the traditional policy solution has been to *internalize the externality*. For example, a polluting firm imposes external costs on others in society, and the

traditional solution has been to internalize the externality by making the firm pay the full social cost of producing the good.

Transferring this idea over to the behavior of an individual agent, an internality—a "within-person externality" (Bhargava and Loewenstein 2015, p. 396)—is the cost to the individual associated with not behaving in a fully rational way. The mistakes that individuals make have costs to the individuals themselves and these costs are internalities. LP-based policies that nudge the agent into more rational action will thus reduce these internalities in precisely the same way that a tax or other environmental regulation would reduce the externality of the polluting firm. As George Loewenstein and Emily Haisley explain:

> Paternalistic policies have the goal of benefiting people on an individual basis … Whereas the conventional justification for government regulation is to limit *externalities*—costs people impose on other people that they don't internalize—to promote the public good, the justification for paternalism is to limit *internalities*—costs that people impose on themselves that they don't internalize … (Loewenstein and Haisley 2008, p. 212)

Returning to Econs and Humans, it seems that Econs are internality-free Humans (or Humans are internality-plagued Econs) and LP nudges are various ways to help Humans eliminate their internalities and behave according to their inner rational agent.

As noted above, one aspect of the existing LP literature is that it often jumps from an analysis of the various parts of the LP argument—Econ, Humans, and such—to particular policies which are so complex that these analytical distinctions can get lost. This section will take a different approach by looking at a model of Econ decision making where the relevant distinctions are clear and straightforward. The model is in many ways an exemplar of *homo economicus*; it is the backbone of twentieth-century microeconomics and played a key role in textbooks and economists' intuition since the 1940s. It is the standard utility-maximizing budget-constrained consumer choice model. Granted most of the discussion of rational choice theory focuses on risky choice and expected utility theory, but consumer choice theory is a simpler case that allows us to exploit the idea of an internality.

In the certainty case, an Econ purchasing a set of goods $x=(x_1, x_2, \ldots x_n)$, facing fixed (competitive) prices $p=(p_1, p_2, \ldots p_n)$ and fixed money income (M) would satisfy his/her true preferences by solving the following, well-defined constrained optimization problem:

$$\begin{aligned} &\text{Max } U(x) \\ &\text{subject to: } \sum_i p_i x_i = M, \end{aligned}$$

where the utility function represents the agent's true preferences. Let's call this the Econ Consumer Choice Problem (ECCP). The solution to ECCP is a set of *n consumer demand functions*:

$$h_i = h_i(p, M) \text{ for all } i = 1, 2, \ldots, n.$$

Econs solve this problem perfectly while Humans have the utility function $U(x)$, but fail to solve the problem correctly; they make mistakes. In the fully optimizing case these demand functions will satisfy certain potentially observable comparative statics conditions[7] and making mistakes means that either the consumer does not have demand functions or their demand functions are missing some of these properties.

Given the ECCP framework, Econ behavior is crystal clear; *Econs will always "be on their demand functions"*; in other words, for any particular vector of prices and money income (p, M), Econs will solve the constrained maximization problem and choose $h_i = h_i(p, M)$ for all $i = 1, 2, \ldots, n$ where each of the h_i functions satisfy the standard restrictions. This is what Econ consumers do. Humans on the other hand, are making mistakes—non-optimal choices—and thus will be *"be off their demand functions."* Of course no consumer in a microeconomics textbook is *ever* off their demand functions since textbooks are concerned with exemplary Econ behavior. So given this, what does a LP nudge do? *They nudge Humans into optimal behavior and thus onto the demand functions that they would have if they were Econs*.

Although thinking about LP as getting Humans back on their demand curves has not been a part of the recent discussions, it is the way that the problem was framed in the original asymmetric paternalism paper (Camerer et al. 2003). They characterized mistakes in terms of internalities and used the analogy of nudges getting individuals back to their optimal demand curves from their mistaken demands.

> When consumers make errors, it is as if they are imposing externalities on themselves because the decisions they make (as reflected by their demand) do not accurately reflect the benefits they derive. The goal of asymmetric paternalism is to help boundedly rational consumers make better decisions and align their demand more closely with the true benefits they derive from consumption. (ibid., p. 1221)

Not only did they frame the nudging problem in terms of being on the fully rational demand curve, they also emphasized, as above, that the problem to be solved by nudging is a mistake (i.e., in the decision-making process) and not the irrationality or instability of the agent's preferences. They stress that not everything that appears to be irrational is irrational (the choice could be what the agent actually prefers like junk food Fred or hermit Sally). The authors use the example of extended warranties. It may be that people make mistakes when they buy such warranties and they do not realize how unlikely such expenses are, but it may be that even fully informed they would still do it (i.e., it is not a mistake for them), they just put a high value on peace of mind. These authors, unlike Thaler and Sunstein who tend to present LP nudging as straightforward, note that such policies must be

[7] There are slightly different characterizations in the literature, but the h_is being homogeneous of degree zero, having negative substitution effects, symmetric cross-partial derivatives with respect to all prices, and a negative semi-definite substitution matrix are the most common (see any advanced microeconomics textbook).

preceded by a careful investigation which sorts out these two possibilities: "in order to properly assess asymmetrically paternalistic policies, we must carefully address whether patterns of apparently irrational behavior are mistakes or expressions of stable preferences" (ibid., p. 1254). Thus, it seems that thinking in terms of internalities and getting back to individual demand curves is not only a useful way to think about LP nudging, it is also an approach that is more likely to inject a note of caution into the discussion of LP.

3 What's it like to be an econ?

Econ behavior is clearly behavior consistent with rational choice theory, but what exactly is rational choice theory? Rational choice has traditionally been seen as a particular version of *instrumental rationality* (using the most appropriate means to achieve given ends) that is constrained in at least three specific ways. First, the ends or goals are *given* and remain stable throughout the analysis. Secondly, the content of the given ends is entirely open. An agent can have the goal of consuming five pounds of chocolate a day and set about to accomplish that goal in a relentlessly rational way. This topic will be discussed in more detail below, but here the point is simply that rational choice theory *alone* does not necessarily imply behavior that coincides with the well-being of the individual agent. Rationality for textbook *homo economicus* is about how goals are pursued, not what the goals are: *de gustibus non est disputandum*.

Thirdly, while the content of preferences is wide open, the structure of those preferences is not. Since preference satisfaction is the goal, preferences must have sufficient structure so that the "most appropriate means" exist. The core structural restrictions on preferences are completeness and transitivity. These are minimal conditions; traditional demand theory, for example, adds restrictions such as convexity and monotonicity so that the resulting demand function is well-behaved. These assumptions will obviously vary from application to application, but the point is that they are restrictions on the *structure* of preferences and not the *content* of preferences. Having intransitive preferences, or having transitive preferences and not acting rationally on them, makes one irrational, while having complete and transitive preferences that are heavily weighted toward candy, fried food, and cigarettes may be perfectly rational (just very unhealthy). As Daniel Hausman and Michael McPherson put it: "People's preferences are rational if they are complete and transitive, and people choose rationally if their choices are determined by their preferences" (2006, p. 60). Econs and Humans both satisfy the first condition, but Humans often fail to satisfy the second.

So far Econs, Humans, and LP policies have been discussed in terms of making individuals more rational—in the certainty case, being on their demand curves—Thaler and Sunstein also say that LP nudging is designed to "make choosers better off, as judged by themselves" (Thaler and Sunstein 2009, p. 5), and making rational choices doesn't necessarily imply being "better off" (even as judged by the agent). For example, people who care about the environment—i.e., prefer a clean environment—often make sacrifices in comfort and lifestyle in order to help protect the

environment, but it seems unlikely that they would say they are individually better off because of such sacrifices. This is just one example of the class of deviations from *homo economicus* Amartya Sen called *commitment* in his famous "Rational Fools" paper (Sen 1977), but many of the anomalies of behavioral economics can cause similar deviations, for example social preferences.[8] The bottom line is that Econs are rational and LP nudging aims at making Humans behave like Econs, but Thaler and Sunstein seem to want more than just rationality, they also want LP nudging to make Humans better off from their own point of view. So how can the textbook *homo economicus* be enhanced so that Econs not only make rational choices, but also make them "better off as judged by themselves"?

There many ways to answer this question, but economists have traditionally answered it by requiring agents to be *self-interested* or *self-regarding*: assuming that Econs prefer x to y if and only if they believe that x is better for them than y. If agents prefer that which they believe makes them better off, then having such self-interested preferences and acting rationally on them would mean that what people prefer is in fact what makes them "better off as judged by themselves." As Hausman and McPherson explain:

> Start with the theory of rationality and add a common assumption of positive economics: that individuals are exclusively self-interested. If nothing but self-interest affects S's preferences, then S prefers x to y if and only if S believe that x is strictly better for S than is y. Rational and exclusively self-interested individuals always prefer that they believe to be better for themselves over what they believe to be worse. (2006, p. 64)

Of course rational choice theory alone does not require self-interest—only completeness and transitivity—but economists have traditionally assumed it. Assuming Econ are self-interested completes the circle from preference satisfaction to being better off as judged by oneself.

So the bottom line for this part of the story is that Econ have preferences which are rational and stable, but also self-interested. If such an agent acts rationally on such preferences they will choose that which they believe will make them better off. Thus, Econ are fully rational and make no mistakes that would motivate or justify (pro-self) nudging.[9] Not only is this characterization of Econ preferences consistent with most of standard microeconomics and much of the philosophical literature on LP, it is also consistent with many characterizations of why pro-self-nudging is

[8] There is an extensive literature on social preferences. See Güth et al. (1982) for an early classic study.

[9] The assumption of self-interest also avoids all the thorny problems associated with altruism and/or malevolence. If A is altruistic toward B but is irrational, then a nudge that makes A more rational will make B better off. But this means that a LP nudge—supposedly purely pro-self—makes a Pareto improvement and also produces a positive externality. But maybe not. Maybe A is altruistic toward B but since this is based only on A's judgments about what would make B better off and may not actually do so. And such complexities go on and on. Real people are altruistic and malevolent and positive rational choice theory often needs to address such issues, but that is not the case for LP which is, as the name suggests, about paternalism and not about third-party effects.

needed: to "counteract cognitive and emotional barriers to genuine self-interest" (Loewenstein and Haisley 2008, p. 215).[10]

Of course accepting this characterization of Econ preferences is not the full story of what it is like to be an Econ. The missing piece—that which Humans lack—is to act rationally on those preferences. To make optimal decisions, the decisions they would have made "if they had paid full attention and possessed complete information, unlimited cognitive ability, and complete self-control" (Thaler and Sunstein 2009, p. 6). But unlike specifying the necessary restrictions on Econ preferences, it is essentially impossible to document what exactly needs to be done to act optimally given those preferences. Mistakes can happen in an infinite number of ways—literally, the consumption of a particular good could be incorrect by 1 unit or 103.765 units—but mistakes are not just about incorrect outcomes, they also involve incorrect beliefs, probabilities, miscalculation (for many reasons), i.e., because of all of the various types of HB mistakes. As a result, the ways that a Human can have the preferences of an Econ but fail to act rationally on those preferences is extremely complex. Of course the number of ways that real humans can go astray is even greater. Real humans might not have preferences that are complete, or transitive, or stable, and in fact they might not have preferences at all. Even assuming that a real human being have preferences, those preferences could be altruistic or malevolent, or involve many other factors that would make their behavior quite different from that of Econ, and yet could not be corrected by LP strategies.

Thus far, we have been discussing the preferences of Econ as *stable*, at least for as long as the relevant period of analysis, in addition to the other conditions such as complete, transitive, and self-interested. However, there is a substantial amount of behavioral literature that suggests that preferences are not (even locally) stable, but rather are constructed in the context of specific choice situations: the extensive behavioral literature on *constructed preferences*.[11] It argues that preference construction is a complex process that is contingent on details of the particular choice situation:

> … the preferences themselves are determined not only by our knowledge, feelings, and memory but also by many aspects of the decision environment, including how the preference objects are described, … The variability in the ways we construct and reconstruct our preferences yields preferences that are labile, inconsistent, subject to factors that we are unaware of, and not always in our own best interests. Indeed … the very notion of a 'true' preference must, in many situations, be rejected. (Lichtenstein and Slovic 2006a, p. 2)

[10] It should be noted that while self-interest solves the "better off as judged by themselves" problem, it is a strong assumption. There is a recent literature that investigates this topic in greater detail and with more emphasis on application (Cartwright and Hight 2019; Sugden 2018; Sunstein 2018 and others).

[11] See Lichtenstein and Slovic (2006b) for an collection of the most important research on constructed preferences. The constructed preference literature originated in the psychological research on *preference reversals* from the early 1970s (Lichtenstein and Slovic 1971).

Constructed preferences are indeed a challenge to rational choice theory and therefore to much of traditional economic analysis, but even if the argument is entirely correct, there is really no place for a discussion of this phenomenon within LP theory or policy. As argued several times above, if Econ is the normative standard and defined by textbook *homo economicus,* then it is a given that Econ has fixed well-behaved preferences—as does the inner rational agent of Humans—and thus neither type of agent could have constructed preferences. Real people may well have constructed preferences—or no preferences whatsoever—but such people are not Econ and they cannot be nudged into being Econ, because there is no coherent way of talking about mistakes in rational decision making unless there are stable well-ordered preferences to serve as the normative reference point. If a Human's preferences were constructed within the context of choice there would be no way to design a nudge that would move them into better satisfaction of their preferences since their preferences would not come into existence until the agent was engaged in the choice process itself. No one can help you correct a mathematical error when the mathematical problem you are trying to solve only comes into existence when you begin the process of solving it, and keeps changing as a result of you working on it. But this argument extends to *any* type preference change, not just constructed preferences.

Given all this, both Econs and Humans have stable rational preferences and it is important to emphasize that those preferences *need to be stable* in at least *four ways*: (1) in the traditional way that economists have assumed stable preferences, i.e., they are not changing with respect to new information, interaction with other agents, etc., (2) preferences are *context independent* (they do not change with the choice context), (3) each agent has a single stable preference order (in particular the agent's preferences do not change as a result of the interactions of multiple selves within the inner rational agent), and (4) preferences are *not constructed* in the act of choice. The mistakes of Humans do not come from having something wrong with their preferences, but rather from their outer psychological shell that leads them to the wrong choices, given their preferences. This is just a result of what it is like to be an Econ, and in turn, what it is like to be a faulty Econ (i.e., a Human). The reference point of stable well-behaved preferences—often called *true*, or *latent*, preferences—is necessary for Econs to play the proper normative roll with respect to the mistakes of Humans. As Sugden notes this "is why Thaler and Sunstein need the concept of latent preference—with all its problems" (Sugden 2018, p. 11). For the rest of this paper, the term "preferences" should be taken to mean these true or latent preferences.

So the conclusion is that while constructed preferences may well be an issue for real people making real decisions, LP's commitment to Econs as the proper normative baseline means that constructed preferences *play no role in LP theory or practice*. Since constructed preferences are often considered to be the most powerful critique that has emerged out of the behavioral economics literature, this means that LP—which supposedly puts behavioral economics to work in a serious way—turns a completely blind eye to one of behavioral economics most challenging insights. And this can also be said about other behavioral departures from *homo economicus* such as social preferences and various types of commitment. It was noted earlier that

when Econ is taken as the starting point, LP seems to be a very weak policy tool, but this shortcoming involves a deeper issue. LP was supposed to bring behavioral insights into policy discussions and yet LP has nothing to say about a world consisting of agents with such preference aberrations.

4 Welfare and related issues

Thus far, LP has been discussed both in terms of getting Humans to behave rationally (*homo economicus*) and in terms of making Humans "better of as judged by themselves," but the LP literature often discusses LP nudges in terms of increasing the welfare or well-being of those being nudged. This raises the question of how welfare is defined and/or measured as well as how increased welfare relates to being more rational and/or being better off. Although welfare is a difficult and controversial topic, there is little controversy about welfare among defenders of LP because the concept of welfare that is implicit in LP is the concept of welfare that has been standard in mainstream economics since the 1940s: the *individual preference satisfaction* (IPS)-based view of welfare "which assesses outcomes, policies, and institutions exclusively by how much they enhance or diminish welfare, as measured by the extent to which preferences are satisfied." (Hausman et al. 2017, p. 147). Adopting the IPS conception of welfare closes the circle on all of the various types of improvements that LP policies are aimed at achieving. If a successful LP policy nudges Humans into more rational decision making, then their preferences will be better satisfied, which means they are better off as judged by themselves and also have higher welfare.

The IPS view of welfare has its origins in eighteenth- and nineteenth-century hedonistic utilitarianism, but differs from utilitarianism in several respects. Perhaps the most significant difference is that hedonistic utilitarianism is a *substantive* theory which provides an account of what welfare *is*—hedonistic *feelings of pleasure and pain*—while individual preference satisfaction is a *formal* theory of welfare that "does not say what things are good for individuals, instead it says how to find out: by seeing what people prefer" (Hausman and McPherson 2006, p. 119). Of course there are many other substantive theories of welfare—John Rawls' "primary goods" (Rawls 1971), the "capabilities" view of Sen (1992), Sugden's "opportunity" approach (Sugden 2004, 2010), views based on various lists of measurable outcomes (life expectancy, infant mortality, etc.), and many others—that challenge the dominant IPS view, but at this point they remain minority positions within economics. Given the fact that one of the goals of this paper has been to employ well-known traditional economic concepts in an attempt to better understand LP, the rest of this discussion will assume the IPS view of welfare/well-being.

Even though rationality, being better off, and having higher welfare have been discussed—as well as the implicit commitment to IPS which equates them—there is still one more outcome that Thaler, Sunstein, and others in the LP literature often claim results from a successful LP policy. It is something that, for want of a better term, would make the agent *really better off*. For example, Thaler and Sunstein say:

> The paternalistic aspect lies in the claim that it is legitimate for choice architects to try to influence people's behavior in order to make their lives longer, healthier, and better. (Thaler and Sunstein 2009, p. 5)

Of course "better" can be translated into increasing preference satisfaction, but living longer and being healthier seem to be very specific measurable outcomes that are not necessarily tied to preference. One argument might be that Thaler and Sunstein have shifted to a substantive conception of welfare, but given the deep dependency of LP on IPS, that seems unlikely.[12] If we are willing to assume that things like longer life and better health really do make people better off, then the most straightforward way to connect rational choice with choosing in such a way as to make oneself really better off, is by assuming *perfect knowledge*.

Employing a slightly modified version of the argument in Hausman and McPherson (2006, pp. 64–65), adding perfect knowledge gives the following relationships:

First Rational Choice Theory:

(R1) Agents have true preferences
(R2) Agents act rationally/optimally/in an instrumentally rational way given those true preferences

So (R1) + (R2) = Rational Choice Theory
Add two additional assumptions:
Self-interest (SI) and Perfect Knowledge (PK):

(SI) Agents prefer x to y iff they believe x is better for them than y
(PK) Agents have perfect knowledge about what does and what does not really make them better off

Now putting (R1), (R2), (SI), and (PK) together we have:

Agents choose what they most prefer and they prefer x to y iff x really makes them better off than y

So this completes the better off-welfare-preference identity. If agents have true preferences and act rationally on them, their choices will be rational. If they are self-interested those choices will reflect what they believe is best for them. So Econs make choices that satisfy (R1), (R2), and (SI). If we add (PK) then the preference/utility-maximizing behavior of an Econ they become something of a super-Econ. Such super-Econs not only have well-ordered preferences and act rationally on them, but under (PK) these choices necessarily make them better off. Such an agent will *never make mistakes in either rationality or in what really makes them better off*;

[12] Another option is simply that Thaler and Sunstein (2009) is a popular, rather than academic, book and employs popular rhetoric that should not be taken seriously. Of course since the purpose of this paper is precisely to examine LP by taking *homo economicus* and IPS seriously, the rhetorical account will not be discussed here. For critical discussion of Thaler and Sunstein's rhetoric, see, for example, Berg and Gigerenzer (2010), Rebonato (2012) and Sugden (2018).

correspondingly they are perfect judges of their own best interest and need no help in decision making. Such super-Econ is the Econ of traditional welfare economics and will not be made better off by either LP nudging or traditional paternalism. As Hausman notes:

> If what Marie chooses is best for her, then it is impossible to make her better off by overriding her choices. Although paternalism is obviously not impossible, economists have been happy to have this way of silencing all questions about paternalism. (Hausman 2018, p. 197)

Now that we have all these assumptions laid out, let's start dropping some. First let's drop (PK); the agent is still rational and self-interested but although they are making choices that rationally satisfy their true preferences, they may not be doing what is really best for them. *This is an Econ*; they are acting in a fully rational way and making no HB mistakes, but they may not be making choices that are really good for them; perhaps they have strong preferences for eating fatty foods or smoking cigarettes. Of course, they *may* be making choices that are really good for them; it is just not necessarily the case. Econs are the choosers in microeconomic textbooks; they may have preferences for things which are harmful, but this is just what it is like to be an Econ.

Finally let's drop (R2); the agent has rational and self-interested preferences but does not choose optimally; they make mistakes in their decision making. This is the Human, the agent whose outer psychological shell is preventing fully rational decision making. This is an agent who could be LP-nudged into behaving more rationally and being better off as judged by themselves.

What all this boils down to is that if we take Econs and Humans seriously, LP nudging is an extremely weak policy tool. It is weak in part because even if it were entirely effective, it only deals with an extremely small set of ways that the behavior of agents could deviate from the rationality of Econ. Perhaps a large portion of human decision making is driven by factors and mechanisms that are not based at all on beliefs, desires, and instrumental rationality. But even if folk-psychological beliefs and desires are behind much of real human decision making—even perhaps consistent desires and epistemically warranted beliefs—the relevant causal mechanisms as well as the outcomes could still be quite different from those of Econ. Perhaps choice is driven by beliefs and desires, but preferences are intransitive, unstable, or constructed. These concerns emphasize the earlier point that successful LP nudging doesn't even correct for many of the important anomalies identified within the behavioral economics literature. In addition, even in the case of a fully equipped Human with well-ordered true preferences and making only HB-based mistakes, successful LP nudging would only make them really better off—"make their lives longer, healthier, and better"—under the heroic assumption of perfect knowledge. Finally, add to the fact that LP nudging is exclusively self-nudging and need not have any direct connection with the traditional concerns that motivate microeconomic-based social policy, and we have a very weak policy tool indeed.

It should be noted that there exists a diverse critical literature on LP which runs parallel to some of the issues discussed in this paper. These concerns are often called *Epistemological* problems since they focus on knowledge and draw on resources

from epistemology, philosophy of science, cognitive psychology, and related fields. A sample of this research includes: Berg and Gigerenzer (2010), Congiu and Moscati (2018), Grüne-Yanoff (2012, 2016), Grüne-Yanoff and Hertwig (2016), Gigerenzer (2015), Guala and Mittone (2015), Hausman (2016), Heilmann (2014), Infante et al. (2016), McQuillin and Sugden (2012), Rebonato (2012), Rizzo and Whitman (2009), Sugden (2008, 2015, 2017, 2018) and Whitman and Rizzo (2015).

One of these epistemological problems that gets a significant amount of attention is in the background of the above discussion and it is useful to draw attention to it. It is what has been called the *interpersonal intelligibility of preferences* problem (Rebonato 2012): the problem that the nudgers/social planners simply cannot know what they would need to know—particularly the agent's true preferences—to design effective LP nudges. As Hausman explains:

> If the object … is to satisfy the … preferences of the inner agent, then economists have to be able to find out what those preferences are … when behavioral economists such as Thaler suggest that cafeteria managers should put the cake in the back, they typically have very little detailed evidence. It seems instead that they believe themselves to be wise third parties, who know that fruit is better for almost everyone and who for that reason attribute a … preference for fruit to most of those served by the cafeteria. But if the object is to satisfy … preferences rather than to provide consumers with what the behavioral economist judges to be best for them, this is a precarious practice. Behavioral economists who believe that they promote well-being by satisfying … preferences need to know what people's … preferences are, not what they should be. (Hausman 2016, p. 28)

Although a very wide array of concerns have been raised in this epistemically focused literature, and some arguments certainly seem stronger than others, it is fair to say that the majority of this research is in general quite *consistent with the account of Econs, Humans, and LP-nudges* provided in this paper. Not only is the account given here consistent with the majority of these criticisms, it also identifies some new concerns such as emphasizing how few of the decision-making errors that are possible would be corrected by LP nudging, as well as how few of the important insights of behavioral economics are actually addressed by LP nudging. It also helps clarify many of the important distinctions that are often blurred within the existing literature, such as: rationality versus being better off, Humans versus real human beings, and pro-self versus pro-social nudges.

Finally, it is important to note that this paper has only been about LP nudges, and in particular, LP nudges that take Econs seriously. Although there were various comments about pro-social nudges in the above discussion, it was always in reference to what they are not—that is, they are not LP nudges—rather than any systematic discussion of what they are, could, or should be. That discussion will not be attempted here, but it is useful to note that nothing said in this paper should be interpreted as a criticism of using nudge-based policies to address social concerns of the traditional microeconomics sort (externalities and public goods) either as new tools or in combination with existing taxes, subsidies, and regulations. And it should be noted that individual nudges may be quite effective with people who have revealed

that they are struggling with certain types of decision making (doing things they would in fact prefer not to do) by say, purchasing things to help them stop smoking, or joining weight watchers, or going to a therapist who addresses such problems. In other words, the account of LP offered here is consistent with recent arguments for a more integrated view of both social policy and individual decision making that includes various types of nudging along with other more traditional policies and solutions (Bhargava and Loewenstein (2015), Guala and Mittone (2015), Loewenstein and Chater (2017) and others).

5 Conclusion and some general remarks

Rather than simply summarizing the various arguments offered in this paper, this last section will respond to some potential criticisms of the paper's Econ-based approach and also try to motivate and/or justify this approach.

Since there are many different interpretations of LP nudging in the literature, some readers may find the austere interpretation of Econs (and Humans) offered here to be unfair to those who support LP. After all, the goal of LP nudging has been characterized quite narrowly, and yet the LP literature replete with stories about nudges that: achieve important social goals (not just satisfy individual preferences), make people substantially better off (not just make them act like *homo economicus*), and benefit a wide range of real human beings (and not just narrowly defined Humans). Shouldn't we pay more attention to the good they are trying to do and pay less attention to the specific things they say about Econs, Humans, and such? Not necessarily and here are some reasons for the approach taken here:

- The argument is not in any way against nudging in general or against innovative new ideas in microeconomic policy. The previous section endorsed the use of nudging techniques to address social issues and made it quite clear that the critical points of the paper were only directed at the way LP was originally characterized by Thaler and Sunstein and not at the idea of nudging in general. While the world might be a better place if nudging techniques were broadly applied to getting people to act more in the public interest or in ways that were actually good for them (whether they prefer it or not), the fact is that this would no longer be *libertarian paternalist* policy.
- Closely associated with the previous point, LP depends heavily on Econ rationality for its uniqueness and originality. In later work Thaler and Sunstein—particularly Sunstein (2016, 2018)—have often sounded like they didn't really mean what they said about Econ being the sole normative standard for LP nudging. But the fact is that using some fairly narrow notion of *homo economicus* as the standard for non-coercive and non-incentive-based paternalist policy *is a key aspect of LP that differentiates it from other forms of paternalism motivated policy*. If it becomes a more generic set of policy tools it disappears as a novel, or even specific, approach to microeconomic policy.
- There is also the argument, noted previously in this paper, that a great amount of clarity comes from starting with something as established as *homo economicus*.

Despite debates about whether an optimizing Econ is the best way to predict or explain individual behavior, every microeconomics textbook contains the same basic optimizing behavior. We may not agree about which particular LP policies will work consistently with which agents, and we may not know how to overcome the various epistemological problems of LP, but we do know what Econs are. Even those sympathetic to various heterodox schools of economic thought, long critical about the scientific adequacy of *homo economicus*, are clear about what Econs are. Given that Thaler and Sunstein explicitly *say* that Econ are the model for correct decision making, and given the amount of talking past each other that seems to go on within the existing LP literature, starting with something as clear as Econ is surely worth a try.

- It is important to note that starting with idealized Econ and ending up being quite critical of the resulting LP policy need not imply a general criticism of the use of rational choice models in economics. It may be quite reasonable to characterize individual behavior in terms of acting optimally on stable well-behaved preferences for certain individuals, in particular contexts, and for certain economic questions and problems; and yet not embrace rational choice theory as the sole normative standard for rational behavior (and thus treating those who live by any other normative standard as being fundamentally faulty and in need of corrective nudging).
- Finally, economists build models for many different purposes (Morgan 2012), but one reason is to strip away the complexity of the situation in order to better identify some of the fundamental relationships and mechanisms in the target domain. In this sense, this paper has explored a particular model of LP. Starting with two key features of Thaler and Sunstein's LP program—(1) a sharp distinction between Econs and Humans, and (2) defining Econs explicitly as *homo economicus*: "the textbook picture of human beings offered by economists"—the paper tried to identify some of the fundamental relationships and implications that are associated with LP nudging, and at various points to even compare it to pro-social nudges, traditional paternalism, and more traditional incentives-based approaches to microeconomic policy.

References

Barton A, Grüne-Yanoff T (2015) From libertarian paternalism to nudging—and beyond. Rev Philos Psychol 6:341–359

Berg N, Gigerenzer G (2010) As-If behavioral economics: neoclassical economics in disguise? Hist Econ Ideas 18:133–166

Bhargava S, Loewenstein G (2015) Behavioral economics and policy 102: beyond nudging. Am Econ Rev 105:396–401

Camerer CF, Loewenstein G (2004) Behavioral economics: past, present, future. In: Camerer CF, Loewenstein G, Rabin M (eds) Advances in behavioral economics. Princeton University Press, New York, pp 3–51

Camerer C, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M (2003) Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. Univ Pa Law Rev 151:1211–1254

Cartwright AC, Hight MA (2019) 'Better Off as judged by themselves': a critical analysis of the conceptual foundations of nudging. Camb J Econ. https://doi.org/10.1093/cje/bez012
Congiu L, Moscati I (2018) Message and environment: a framework for nudges and choice architecture. Behav Public Policy. https://doi.org/10.1017/bpp.2018.29
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Dhami S (2016) The foundations of behavioral economic analysis. Oxford University Press, Oxford
Gigerenzer G (2008) Rationality for mortals: how people cope with uncertainty. Oxford University Press, New York
Gigerenzer G (2015) On the supposed evidence for libertarian paternalism. Rev Philos Psychol 6:361–383
Gigerenzer G, Brighton H (2009) *Homo Heristicus*: why biased minds make better inferences. Top Cognit Sci 1:107–143
Grüne-Yanoff T (2012) Old wine in new casks: libertarian paternalism still violates liberal principles. Soc Choice Welf 38:635–645
Grüne-Yanoff T (2016) Why behavioural policy needs mechanistic evidence. Econ Philos 32:463–483
Grüne-Yanoff T, Hertwig R (2016) Nudge versus boost: how coherent are policy and theory? Mind Mach 26:149–183
Guala F, Mittone L (2015) A political justification of nudging. Rev Philos Psychol 6:385–395
Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. J Econ Behav Organ 3:367–388
Hagman W, Andersson D, Västfjäll D, Tinghög G (2015) Public views on policies involving nudges. Rev Philos Psychol 6:439–453
Hansen PG (2016) "The definition of nudge and libertarian paternalism: does the hand fit the glove? Eur J Risk Regul 7:155–174
Hausman DM (2016) On the econ within. J Econ Methodol 23:26–32
Hausman DM (2018) Philosophy of economics: a retrospective reflection. Rev Econ Philos 18:183–201
Hausman DM, McPherson M (2006) Economic analysis, moral philosophy, and public policy, 2nd edn. Cambridge University Press, Cambridge
Hausman D, McPherson M, Satz D (2017) Economic analysis, moral philosophy, and public policy, 3rd edn. Cambridge University Press, Cambridge
Heilmann C (2014) Success conditions for nudges: a Methodological Critique of Libertarian Paternalism. Eur J Philos Sci 4:75–94
Herrnstein RJ, Loewenstein GF, Prelec D, Vaughn WJ (1993) Utility maximization and melioration: internalities in individual choice. J Behav Decis Mak 6:149–185
Infante G, Lecouteux G, Sugden R (2016) Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. J Econ Methodol 23:1–25
Kahneman D (2003) Maps of bounded rationality: a perspective on intuitive judgment. Am Econ Rev 93:1449–1475
Kahneman D, Tversky A (1979) Prospect theory: an analysis of decisions under risk. Econometrica 47:263–291
Kahneman D, Tversky A (eds) (2000) Choices, values, and frames. Cambridge University Press, Cambridge
Kahneman D, Knetsch JL, Thaler R (1991) Anomalies: the endowment effect, loss aversion, and status quo bias. J Econ Perspect 5:193–206
Lichtenstein S, Slovic P (1971) Reversals of preference between bids and choices in gambling decisions. J Exp Psychol 89:46–55
Lichtenstein S, Slovic P (2006a) The construction of preference: an overview. In: Lichtenstein S, Slovic P (eds) The construction of preference. Cambridge University Press, Cambridge, pp 1–40
Lichtenstein S, Slovic P (eds) (2006b) The construction of preference. Cambridge University Press, Cambridge
Loewenstein G, Chater N (2017) Putting nudges in perspective. Behav Public Policy 1:26–53
Loewenstein G, Haisley E (2008) The economist as therapist: methodological ramifications of 'light' paternalism. In: Caplin A, Schotter A (eds) The foundations of positive and normative economics: a handbook. Oxford University Press, Oxford, pp 210–245
McQuillin B, Sugden R (2012) Reconciling the normative and behavioural economics: the problems to be solved. Soc Choice Welf 38:553–567
Mongin P, Cozic M (2018) Rethinking nudge: not one but three concepts. Behav Public Policy 2:107–124
Morgan MS (2012) The world in the model: how economists work and think. Cambridge University Press, Cambridge

Rawls J (1971) A theory of justice. Harvard University Press, Cambridge
Rebonato R (2012) Taking liberties: a critical examination of libertarian paternalism. Palgrave Macmillan, New York
Rizzo M, Whitman DG (2009) The knowledge problem in the new paternalism. Brigh Young Univ Law Rev 2009(4):905–968
Schmidt AT (2019) Getting real on rationality—behavioral science, nudging, and public policy. Ethics 129:511–543
Sen A (1977) Rational fools: a critique of the behavioral foundations of economic theory. Philos Public Aff 6:317–344
Sen A (1992) Inequality reexamined. Harvard University Press, Cambridge
Sugden R (2004) The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. Am Econ Rev 94:1014–1033
Sugden R (2008) Why incoherent preferences do not justify paternalism. Const Polit Econ 19:226–248
Sugden R (2010) Opportunity as mutual advantage. Econ Philos 26:47–68
Sugden R (2015) Looking for a psychology for the inner rational agent. Soc Theory Pract 41:579–598
Sugden R (2017) Do people really want to be nudged towards healthy lifestyles? Int Rev Econ 64:113–123
Sugden R (2018) 'Better off, as judged by themselves': a reply to cass sunstein. Int Rev Econ 65:9–13
Sunstein CR (2013) The storrs lectures: behavioral economics and paternalism. Yale Law J 122:1826–1898
Sunstein CR (2015) Nudges, agency, and abstraction: a reply to critics. Rev Philos Psychol 6:511–529
Sunstein CR (2016) People prefer system 2 nudges (kind of). Duke Law J 66:121–168
Sunstein CR (2018) 'Better off, as judged by themselves': a comment on evaluating nudges. Int Rev Econ 65:1–8
Sunstein CR, Thaler RH (2003) Libertarian paternalism is not an oxymoron. Univ Chic Law Rev 70:1159–1202
Thaler RH (1980) Toward a positive theory of consumer choice. J Econ Behav Organ 1:39–60
Thaler RH (2000) From homo economicus to homo sapiens. J Econ Perspect 14:133–141
Thaler RH (2017) Behavioral economics. J Polit Econ 125:1799–1805
Thaler RH (2018) From cashews to nudges: the evolution of behavioral economics. Am Econ Rev 108:1265–1287
Thaler RH, Sunstein CR (2003) Behavioral economics, public policy, and paternalism. Am Econ Rev 93:175–179
Thaler RH, Sunstein CR (2009) Nudge: improving decisions about health, wealth and happiness. Penguin, London
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Whitman DG, Rizzo MJ (2015) The problematic welfare standards of behavioral paternalism. Rev Philos Psychol 6:409–425