Abstract: A little more than a decade ago, Nathan Berg asked the question why behavioral economics had not yet inspired a reorientation of normative economics. Given the ever rising interest in the new, psychologically informed economics, the present paper will outline the way reasoning about the normative implications of behavioral economics has developed since Berg raised the issue. We argue that behavioral economics has inspired new thinking about the prospect of – what we would call – a Behavioral Normative Economics. The paper sketches important approaches in the field, questions their internal theoretical coherence, and outlines initial ideas on how to alter the notion of welfare in normative economics.

Keywords: Behavioral; Identity; Methodology; Normative; Welfare

JEL: B4; B5; B0; A1
1. Introduction

Given that economics has now rediscovered the – often intricate and surprising – ways real-world agents think and choose, it’s time to consider what all the positive insights from Behavioral Economics imply in the way of policy implications. It seems obvious that economists’ policy advice can only benefit from a more realistic underlying model of human conduct. In this brief exposé, we will try to clarify whether this is indeed the case. Since any attempt to develop policy implications in light of insights from positive economics presupposes normative premisses – which should ideally be made explicit – this endeavor is mostly a critical overview of existing approaches to develop what we will refer to as ‘Behavioral Normative Economics’ (henceforth BNE).\(^1\) BNE is the set of all attempts to modify standard normative economics (which essentially comprises ‘welfare economics’ and normative constitutional economics) so that it is better aligned – or ‘reconciled’ (McQuillin/Sugden 2012a) – with insights from behavioral economics and can be coherently applied in a behavioral world, i.e. a world where individuals have limited mental resources.\(^2\)

As we will try to show, in the future the key question shaping BNE – and behavioral policy advice derived from it – concerns the underlying conception of identity: Do real-world human beings consist of ‘multiple selves’ (as presumed in most contemporary contributions to the literature on BNE), or should we rather try to overcome this modeling approach? Obviously, questions like this suggest a need to rekindle the dialogue between (behavioral) economists and ethicists.

We will start with a brief explanation of why we think that economists in general and behavioral economists in particular should be interested in normative theorizing (section 2). Then we explore recent attempts to elaborate upon a purportedly ‘pragmatic’ way to incorporate behavioral economics findings into policy advice (section 3). Given the shortcomings of such attempts, we will then proceed to discuss the branch of BNE that is currently most popular within economics, to wit, ‘Behavioral Welfare Economics’ (section 4). One particularly controversial offspring of that approach has become known as ‘Libertarian

---

\(^1\) We do not claim any copyright for the term. For instance, Hargreaves Heap (2013) mentions “behavioural normative economics” (ibid.: 986) or a “behaviourally informed welfare economics” (ibid.: 992, 996). Hands (2013) talks about a “behavioral economics-based welfare economics” and observes a “new welfare economics”, but concedes that latter term to be misleading since there already exists an orthodox “new welfare economics”, introduced by Samuelson, Bergson and others in the 1940s (ibid.: 22).

\(^2\) By limited mental resources we mean limited computational capacities, attention, and willpower.
Paternalism’ and will be briefly discussed in sub-section 4.2. Section 5 discusses a radical alternative approach to BNE: Robert Sugden’s ‘opportunity’ criterion. Finally, section 6 identifies some promising avenues for future research in the field of BNE.3

2. The significance of behavioral arguments in normative reasoning

A little more than a decade ago, Nathan Berg posed the question, why behavioral economics seemed to have almost no impact on “discussions about how economic policy ought to be made” (Berg 2003: 411). Put differently, why was the ascent of behavioral economics – already clearly visible then4 – not accompanied by a “new normative framework for analyzing policy” (ibid.: 412)? Given the ever rising interest in the new, psychologically informed economics and the practical applicability of its key insights, this question is perhaps even more important today than it was back in 2003.

Berg wrote about “Normative Behavioral Economics”, which, to us, seems somewhat unfortunate. For instance, it might be read as implying that behavioral economics constitutes its own isolated subfield within economics that happens to have, as a sub-subfield of its own, a branch which studies its normative implications. Thus, such a phrasing may seriously understate the relevance of those normative implications for economists in general. What’s more, it also seems to beg the question: Does behavioral economics indeed have normative implications of its own? We deem it more adequate to do some reshuffling: Behavioral Normative Economics suggests that behavioral economics, rather than denoting an independent subfield within economics, pervades, as a method, the whole body of economic theorizing;5 that it makes sense, therefore, to study not only how, say, Labor Economics or

3 We are aware of the fact that we excluded (at least) one very influential field of normative economics, namely happiness research, from our discussion. Happiness economics has clearly inspired new ways of thinking about how to conceptualize and measure welfare, especially in the discussion about gross national happiness index (see, e.g., Frey and Stutzer 2010 or Kahneman et al. 2004). In spite of its general impact on economics and policy making, we believe that its methodological approach is not explicitly behavioral and therefore can be excluded from a survey discussion of BNE. Furthermore, we would argue that the road of promoting happiness as the normative foundation of welfare might better not be taken: normative happiness economics favors a mental-state account to well-being and thereby exhibits inevitable measurement problems, e.g. problem of hedonic adaption, or selective memory and projection biases, as well as serious philosophical flaws, e.g. welfare contains more than just hedonic aspects of pleasure (see Schubert 2012 for a critique of “happiness economics”).

4 Kahneman and Smith had just been awarded the Sveriges Riksbank Prize in Economics in memory of Alfred Nobel.

5 We are inspired here by Simon (1987: 221) who argues that behavioral economics “is best characterized not as a single specific theory but as a commitment to empirical testing of the neoclassical assumptions of human behavior and to modifying economic theory on the basis of what is found in the testing process.”
Law & Economics or even Macroeconomics could benefit from a psychologically informed perspective (turning them into “Behavioral Labor Economics” etc.), but also to probe whether Normative Economics as a whole might be enriched by adopting such a fresh perspective. In our view, such a refocusing greatly enhances the relevance of the issues we will discuss in the present paper, building on Berg’s pioneering 2003 study.

Why should economists in general and behavioral economists in particular be interested in Behavioral Normative Economics (henceforth BNE)? We take it to be near-obvious that economist’s policy advice could benefit from improving the descriptive psychological adequacy of the underlying models. For those who do not share this conviction, the following two considerations may be relevant: First, behavioral economics may ultimately turn out to have a larger impact on normative reasoning than on positive theorizing (Hargreaves Heap 2013: 992-995). For despite all the enthusiasm accompanying their research, its impact on the former may be constrained by two factors: First, its predominant focus is (still) on studying the determinants of behavior at the level of the individual agent, while economics in general has always been interested, first and foremost, in analyzing the – unintended – aggregate effects of behavior, including those effects’ repercussions on individual action. In order to study those, however, more research on institutions might be necessary (e.g. Lewin 1996: 1320). Second, there will always be staunch defenders of the neoclassical approach who resort to the well-known “as-if” argument, popularized by Friedman (1953), which says that competitive pressures make sure that agents will act “as if” following the maxims of rational choice, and hence economists need not bother to study the psychology of human decision-making in order to make their assumptions descriptively adequate. Utility maximization would then not be understood as a description of actual mental processes (how an agent thinks), but as a “representation” of how an individual behaves (e.g. when responding to incentives). To be sure, that argument is flawed on many dimensions, but it may be sufficiently popular to seriously impair the influence of behavioral economics in the realm of positive theory development.

Given these considerations, the impact of behavioral economics on positive theorizing may be limited eventually. On the other hand, its impact on normative reasoning is hardly

---

6 We do not claim to have invented the term. For instance, Hargreaves Heap (2013) mentions “behavioral normative economics” (ibid.: 986) or a “behaviorally informed welfare economics” (ibid.: 992, 996). Hands (2013) talks about a “behavioral economics-based welfare economics” and observes a “new welfare economics”, but concedes that latter term to be misleading since there already exists an orthodox “new welfare economics”, introduced by Samuelson, Bergson and others in the 1940s (ibid.: 22).
affected by the factors adduced above: For normative economics, as the theoretical study of welfare-improving outcomes, it matters a lot what real-world (“behavioral”) agents consider to count as well-being. Friedman’s “as-if” reasoning does not apply to normative economics. What matters is what individuals perceive as welfare-improving, and it’s quite plausible that this perception is influenced by the way information is processed in the human brain. More importantly, orthodox policy implications that are critically based on the assumption that people possess well-behaved preference orderings obviously cannot be applied to a setting where people in fact have incoherent preferences (Hargreaves Heap 2013: 992, 997; see also Sugden 2004, 2008).

The second argument for the relevance of BNE consideration concerns the value of methodological diversity: As Berg pointed out in his paper, standard neoclassical normative economics “systematically guides policy analysis in certain directions rather than others” (Berg 2003: 415). Behavioral models open up a distinct range of policy conclusions. Understanding this hidden normative dimension of methodological choice certainly improves the quality of economic policy advice by minimizing the risk of normative biases.

These arguments notwithstanding, Berg (2003: 412f.) observed that leading behavioral economists – such as Thaler (1991) – not only refrained from drawing implications from their work for normative economics, but even warned their readers not to engage in that endeavor. Thaler (1991: 138) states that the “[…] demonstration that human choices often violate the axioms of rationality does not necessarily imply any criticism of the axioms of rational choice as a normative ideal. Rather, the research is simply intended to show that for descriptive purposes, alternative models are sometimes necessary.” Thaler (1991) advocates to keep the homo economicus firmly in place as a normative role model, presumably in order to make behavioral economics as a whole more digestible for mainstream economists. A related reason may be the worry that, with rationality widely interpreted as a refutable assumption among behavioral economists (as opposed to a metaphysical principle), and with it now being effectively refuted, sticking to rationality in the normative sense seems the only way left to

---

7 This holds as long as normative economics sticks to an individualistic perspective, which we will assume throughout.

8 As Berg (2013: 416) puts it, “in averaging over imperfectly correlated errors (one for each method... considered), the composite tool of analysis (the ‘portfolio’...) is more precise than a single tool is.”

9 See also Angner (2014) for an elaborated analysis why behavioral economists tend to favor neoclassical assumptions about rationality as a normative benchmark.
keep this key notion on board in economics.\textsuperscript{10} In the rest of this exposé, we will explore whether that research strategy worked out well; more generally, we want to see whether progress has been made with respect to the construction of a behaviorally inspired normative economics.

3. Behavioral policy advice – the ‘pragmatic’ way

From the viewpoint of a representative practicing economist, the most straightforward way to incorporate behavioral findings into public policy advice is what we refer to as the ‘pragmatic’ way: Positive insights into, say, the way real humans process monetary incentives in experiments, are built into instrumental reasoning, without questioning the underlying normative premisses. In other words, some policy goal is taken as given, and then the ‘best’ (read: most efficient) means to realize that goal are investigated, taking the state of the art of behavioral science into account.\textsuperscript{11} Prominent recent advocates of this approach are Bolton and Ockenfels (2012) – who refer to it as ‘behavioral economic engineering’ – Chetty (2015), and Madrian (2014).\textsuperscript{12}

To illustrate, if you are a kindergarten director wishing to reduce the number of parents who are late for picking up their children in the afternoon, it may be a good idea not to introduce a monetary fee, for that has been shown to crowd out certain social norms, generating counterproductive behavioral effects (Gneezy and Rustichini 2000). In this and many other contexts, standard economic analysis fails to adequately grasp the source of impediments to market efficiency. Psychological and cognitive biases of consumers, and not

\textsuperscript{10} Hands (2013) notes that, with the descriptive interpretation of “rationality” losing ground, its normative interpretation seems to become more and more popular.

\textsuperscript{11} Following John Neville Keynes’ (1917) famous distinction, the authors favor an instrumental understanding of economics (in Keynes’ terminology: “the art of political economy”) as opposed to normative economic theorizing (in Keynes’ terminology: “applied ethics”). While the latter examines alternative policy-goals and the suitability of welfare criteria, the former only assesses institutional policies in the light of given ends.

\textsuperscript{12} See also Bhargava and Loewenstein (2015).
their rational behavior, are assumed to be a main driver for ‘behavioral market inefficiencies’ (Madrian 2014: 681).13

The engineering approach considers neoclassic incentive regulation and behaviorally inspired interventions as complementarities and not substitutes. Chetty states that “the decision to include behavioral factors in economic models should be viewed as a pragmatic rather than philosophical choice” (2015: 28). Therefore, theoretical analysis in behavioral economic engineering adjusts standard economic models only at the margin (e.g., modelling boundedly rational agents equipped with social preferences).

Generalizing, this way to incorporate behavioral findings aims at making economic institutions more robust with respect to the presupposed level of ‘rationality’ (in the standard economic sense of consistency) or the institutional context of the agents affected. Behaviorally enriched models can inform the effectiveness of standard economic policies and contribute to the development of a new set of policy tools that are more cost-effective compared to traditional ones (Madrian 2014: 663). Importantly, this approach also puts emphasis on gathering experimental evidence. As Bolton and Ockenfels (2012: 667) put it, “the objective behind behavioral economic engineering is to catch problems in the robustness of new economic institutions and mechanisms prior to their engagement on a large social scale, where failure can be far costlier.”

While this pragmatic approach looks attractive at first sight, it’s important to realize its limits. It essentially leaves open a key question Where do the policy goals come from? And is it the economist’s job to critically discuss those goals as well? We conjecture that a majority of economists would probably rather follow Gul and Pesendorfer (2007) and shy away from venturing into the intricate terrain of questioning policy goals. Depending on your own answer to the second question, above, there are basically three ways to go forward.

First, you may use the ‘behavioral economic engineering’ toolbox merely to hypothetically investigate the effectiveness of alternative instruments to achieve any exogenously given goal. Typically, that goal may be ‘imported’ from the policy-makers’ realm and make more or (rather) less economic sense. This approach will need to rely on some conception of ‘welfare’ in order to be able to compare alternative instruments as to their

13 Madrian (2014) names present bias as an example of a behavioral anomaly that impedes market efficiency. She states that “the tendency of individuals to place much less weight on the future relative to the present than would be predicted by standard models of time discounting” (664) can lead individuals to make decisions today that reduce future welfare in ways that individuals will later regret. In doing so, Madrian reveals the market efficient way is based on individual’s following neoclassical rationality axioms.
‘welfare effects’, and it will typically resort to the notion of welfare developed by standard welfare economics, i.e., welfare as the degree of satisfaction of given and perfectly consistent preferences. As we will see below, this limits the range of issues in which this approach can be coherently applied.

Second, you may find that a critical discussion of policy goals is part of the economist’s job description and argue that policy should aim at those goals that maximize ‘welfare’ in the sense of standard welfare economics (see above). The same caveat regarding the range of applications applies, of course.

Third, you may be skeptical about standard welfare economics and rather argue that policy should support agents’ access to well-functioning markets, rather than maximize allocative efficiency. As Becker (1962) showed long ago, the former requires much less in the way of individual rationality than the latter.\textsuperscript{14}

4. ‘Behavioral Welfare Economics’: Laundering preferences

Assume, for the sake of the argument, that it’s indeed the economist’s legitimate job to critically discuss alternative policy goals. In that case, the question of BNE enters the stage, and we face the challenge of reconciling normative economics with behavioral economics. Roughly speaking, there are three ways to do that: We can either try to save as much as possible of standard normative (or rather: ‘welfare’) economics with its foundation in the revealed preference approach; or we can abandon the traditional focus of normative economics on welfare and redirect our attention to some freedom criterion; or we try to take a step back and focus on the way real-world individuals go about forming their preferences (read: their identity or character) in the first place. In this and the following sections 5 and 6, we will briefly review these three alternative strategies.

4.1 The basic idea

Acknowledging the fact that the individual preferences of real-world human beings are often far from exhibiting the ‘well-behaved’ coherence assumed in the textbook, Bernheim and Rangel (2007, 2008, 2009) have suggested to identify, among the incoherent set of an agent’s

\textsuperscript{14} See also Sugden (2004).
preferences (read: choices), the subset whose satisfaction plausibly indicates a gain in well-being for that very agent.\(^{15}\) Within the confines of that subset, observed choices are respected, which, according to the authors, establishes sound classic-liberal credentials: “Choices provide appropriate guidance because they are choices, not because they reflect something else” (ibid.: 52).\(^{16}\)

The problem, of course, is that the behavioral welfare economist first has to identify the subset of preferences whose satisfaction is then taken to reliably increase an agent’s well-being. In other words, the messy set of actual (‘manifest’) preferences has to be ‘laundered’ first in order to reconstruct what the agent ‘really’ prefers – or what she should prefer, were she free from all those cognitive biases that make her preferences so inconsistent in the first place. Technically, Bernheim and Rangel resort to a multiple-selves methodology that allows them to model the observed choices as maximization under constraints, with a twist: The ‘constraints’ include all sorts of contextual factors – such as framing effects – that are deemed irrelevant in the economics textbook. Hence, a given agent activates different utility functions (i.e., different ‘selves’), depending on the specific situational context he finds himself in, which may result in different choices even when option set, relative prices and budget constraint remain unchanged. Importantly, some good A is only deemed better than some alternative good B if A is preferred over B by all the agent’s ‘selves’.\(^{17}\) Put differently, revealed preferences are being respected as normatively relevant if and only if they are robust under all conceivable contextual constraints and have the standard well-defined structure. In all the other cases – i.e., in all cases that actually give rise to the reconciliation problem – the behavioral welfare economist remains silent. Hence, rather than providing a constructive

\(^{15}\) For a related approach, see Rubinstein and Salant (2012); Manzini and Mariotti (2014) provide a very critical discussion.

\(^{16}\) Bernheim and Rangel (2008) give two arguments in support of this choice-centered approach. First, taking choice as the normative welfare criterion is a commitment to individual autonomy as a good in itself. Secondly, despite behavioral anomalies, revealed choices rank highest among all the deficient welfare measures to detect individual well-being.

\(^{17}\) Similar to the Pareto principle, Bernheim and Rangel develop an intrapersonal welfare criterion for ‘good choices’ which is based on preferences that are ranked above all others revealed by all selves of one person over time. So based on a binary individual welfare relation \(P^*\) and two available options \(x\) and \(y\), their welfare criterion can be formally defined as follows: \(xP^*y\) iff \(y\) is never chosen by any self of the same person within the welfare-relevant choice domain when both \(x\) and \(y\) are available. In such a situation, the person’s preferences over \(x\) and \(y\) most likely turn out to be insensitive to any of the usual behavioral biases.
answer to the questions raised by the reconciliation problem, Behavioral Welfare Economics seems to sidestep the problem.\(^{18}\)

Furthermore, this approach only ranks states when the ‘intrapersonal Pareto condition’ is satisfied. However, it is not immediately clear why, from a normative point of view, one should not have different preference rankings at different points in time. A person might change her lifestyle, thus producing incoherence between the different preference rankings when judged from a ‘unitary self’. But why should we assume such a unitary meta-perspective rather than an evolutionary view of the self (Hargreaves Heap 2013)?

4.2 An application: Libertarian Paternalism

One prominent practical application of the notion of ‘laundering’ manifests preferences in order to uncover or reconstruct the allegedly ‘true’ ones has been suggested by Thaler and Sunstein (2003): Libertarian Paternalism (henceforth LP), a normative public policy program, aims at improving people’s welfare without interfering with their choice sets.\(^{19}\) It does so through the use of nudges, i.e., subtle modifications of people’s situational context (their ‘choice architecture’ in LP lingo) that basically work by exploiting cognitive biases or by responding to them. The ‘nudge agenda’ has become immensely influential among policymakers, particularly in the Anglo-Saxon world.\(^{20}\)

Libertarian Paternalism is too well-known to warrant a detailed description and critique in this brief exposé.\(^{21}\) The point to emphasize in our present context is that its advocates follow (rather implicitly) the general strategy pursued by Behavioral Welfare Economics, viz., to ‘launder’ people’s actual preferences before letting them enter their

\(^{18}\) Whitman and Rizzo (2015: 16) correctly point out that Bernheim and Rangel implicitly make the assumption that “each person has a neoclassical agent deep within himself which is struggling to surface. Decision-making processes are thus deemed to be malfunctioning insofar as they fail to produce choices consistent with the standard preference structure. In other words, malfunction is not independently defined; it is whatever does not make standard choice theory descriptively accurate. This approach thus assumes away the possibility of individuals who simply lack true preferences satisfying the neoclassical axioms of transitivity and completeness.”

\(^{19}\) See also Thaler and Sunstein (2008).

\(^{20}\) The term Libertarian Paternalism is ill-conceived in that there are good reasons to question its 'libertarian' character (e.g. Grüne-Yanoff 2012); beyond that, only a subset of all the applications suggested in books such as Thaler and Sunstein (2008) and Sunstein (2014) are actually paternalistic in nature (Hausman and Welch 2010, Guala and Mittone 2015). Some of them aim at internalizing externalities.

individual (or social) welfare calculus. Sunstein and Thaler (2003: 1162) are explicit in identifying those choices as ‘inferior’ (and in need of correction) that “[people] would change if they had complete information, unlimited cognitive abilities, and no lack of self-control.” The underlying idea is, of course, that Thaler and Sunstein are explicit in recommending homo economicus as the normative role model that informs their whole public policy agenda.\textsuperscript{22}

Without entering the debate about the appropriate use of economic ‘rationality’ conceptions, we merely wish to point to the pitfalls of taking homo economicus as the normative benchmark real-world people are supposed to follow. In the parlance of dual-process theory which is hugely popular among behavioral economists, this is tantamount to privileging an agent’s purportedly ‘long-term’ or ‘System 2’ or ‘planner’ self (with his specific preferences), as opposed to the ‘short-term’ preferences of the spontaneous and affect-driven ‘System 1’ self. It is far from clear why exactly one should accept this approach, given the intuition that people’s allegedly ‘long-term’ preferences (“I wish I didn’t want to smoke!”) are often expressive in nature (Schnellenbach 2012) and given the controversial nature of dual-process theorizing in general (Buturovic and Tasic 2015).

5. ‘Freedom as opportunity’

In a series of papers Robert Sugden (2004, 2008, 2010) has elaborated upon a normative criterion – explicitly couched in a contractarian framework\textsuperscript{23} – that is supposed to show that

\textsuperscript{22} On this, see also Thaler (2015: 25, 29, 57). See also the critique advanced by Gigerenzer (2015), drawing inspiration from the ‘old’ behavioral economics school (Herbert Simon, Vernon Smith etc.).

\textsuperscript{23} Rather than seeing themselves in the role of policy advisors that assess the value of social states from some allegedly neutral and encompassing (“synoptic”) perspective, normative economists should take a rule-based approach, addressing the citizens affected by their advice directly: The task would then be to help negotiating “a fair agreement between individuals, each of whom is looking at the world from his own viewpoint” (Sugden 2008: 229-230).
behavioral economics (the positive findings of which he accepts) is not only incompatible with paternalistic policy implications (in the sense of LP), but positively calls for an anti-paternalistic stance on public policy design.\(^2\) Again, we aim not to present his highly original approach in detail, but will rather focus on what we see as its key contribution to the fundamental debate shaping BNE, namely, the issue of identity (see above).

Sugden’s underlying normative intuition is that every agent should be free to try to satisfy whatever incoherent and possibly eccentric preferences she finds herself with at any given point in time – provided she does not harm other people in the process. While individual preferences may very well be ‘unconsidered’, people’s “valuing the opportunity to satisfy them is considered’ (Sugden 2006: 217, italics added) and can, thus, provide the foundation of an alternative approach to normative economics.\(^2\) Hence, he suggests to take opportunities, rather than specific outcomes or consumption bundles, as the ultimate carriers of value. His normative role model is not homo economicus but rather the ‘responsible agent’ who treats her past, present and future preferences as fully her own, even when facing uncertainty or regret (Sugden 2004: 1018). Sugden claims that his approach overcomes the multiple-selves model of man in that he understands agents as continuing loci of responsibility: They endorse any preference their own selves once had or will have. In light of this role model, it’s always unconditionally good for each agent to maximize her own opportunity set. By pointing to “opportunities” instead of cleansed or nudged preference sets, Sugden admits the possibility of incoherent preferences, while retaining the substance of liberal welfare economics – i.e., the principle of consumer sovereignty.

This approach paves the way for a justification of anti-paternalism that does not narrowly focus on allocative efficiency in the economics textbook sense. Rather, Sugden argues that by providing each participant maximum chances for engaging in mutually beneficial transactions, markets give each agent, “rational or irrational, what she wants and is willing to pay for, when she wants it and is willing to pay for it” (McQuillin and Sugden 2012b: 630f.). The key is that markets give agents incentives to try to predict (and satisfy) what other agents will prefer in future periods (Sugden 2008: 244).

\(^2\) A behavioral-economic case against government paternalism can of course also be made on non-normative grounds, e.g. from a ‘behavioral political economy’ perspective (Schnellenbach and Schubert 2015). See also Glaeser (2006).

\(^2\) Sugden (2010: 49) defines an opportunity for a given agent as “something that he has the power to bring about, if he so chooses”.

At first sight, this approach looks much more attractive than its main counterpart, Behavioral Welfare Economics (with its prominent offspring, Libertarian Paternalism), in that it seems to avoid the – ultimately futile – search for people’s ‘true’ preferences. On the other hand, one may argue (and we do, in fact) that Sugden implicitly still sticks to the dual-process approach that, as we have seen, is so highly influential in BNE. Instead of privileging ‘long-term’ preferences a homo economicus would indulge, he just advocates the very opposite: Agents should be free to pursue their short-term preferences, quite independent of any concern as to the consequences in terms of well-being.

In our view, the critical point concerns a key implication Sugden himself derives from his normative role model (the ‘responsible agent’): He claims that while such an agent may wish to steer the future development of his own preferences by mustering willpower (which he refers to as ‘self-command’), he will never wish to constrain his opportunity set by external means (‘self-constraint’). If this really does follow from the ‘responsible agent’ conception (and if the latter is necessarily involved in the opportunity criterion), then the model of identity underlying this approach is questionable. Most real-world agents, or so we presume, are not bent on maximizing their own personal opportunity sets independent of any concern for its consequences in terms of personal well-being, but rather wish to engage in both self-command and self-constraint (as defined by Sugden), depending on the circumstances. It seems that Sugden ultimately confuses legitimate self-commitment (which would comprise both self-command and self-constraint) with illegitimate paternalism.

---

26 See also Fumagalli (2013) on the pitfalls of looking for ‘true’ utility functions.

27 Sugden (2008: 243) asserts that only by giving authority to the “acting” or “impulsive self”, rather than the self as “maker of plans or source of reflective judgement about the well-being of the continuing person” might one be able to appreciate the value of the market in providing for opportunity. This view implicitly assumes that the mere absence of interferences (i.e. negative liberty) is sufficient to give rise to the quality of agency necessary for the normative role model of the “responsible” person to be empirically plausible. This, however, is at odds with basic insights from behavioral economics (see Schubert 2015a).

28 Sugden (personal communication, March 2013). An example of the latter would be an alcoholic’s strategy to lock away the spirits in a cupboard and throw away the only key (ibid.). On self-command, see also Sugden (2004: 1018). See Schubert (2015a) for a more detailed critique.

29 We are actually unsure about the first conditional.

30 See also Vanberg (2014: 338-341).

As we have argued, the problem of how to conceptualize people’s identity is one (perhaps the) key question whose answer shapes the future development of BNE: Do agents whose choices are marred by cognitive biases consist of multiple selves among which one – the cool ‘planner’ or the hot, spontaneous ‘doer’ – needs to be privileged? Or is there some other way to approach this problem?

Here’s an idea. Given what behavioral economics tells us about the nature of human preferences, it seems that most of them tend to be highly contingent, sometimes even critically dependent on changes in contextual factors that can only be deemed arbitrary. Moreover, processes of preference formation are often not subject to control by the agent herself. Nevertheless, real-world human beings value the fact that they are able to form their own preferences in a way that at least ‘feels’ somehow self-determined (Deci and Ryan 2000).

The highly contingent, yet obviously value-bearing nature of individual preferences leads us to suggest that we should perhaps take a step back and focus on the process of preference formation (or ‘construction’), rather than on single preferences at any particular moment in time. Consider this quote from an earlier debate, in the pages of the American Economic Review, on ‘Consumer sovereignty’:

“Consumers’ choices may not reflect their true tastes; but … maybe these tastes cannot accurately be known; or … they are not really ‘owned’ but only ‘loaned’ tastes anyway, passed on from one person to another. What really can belong to the self and be accurately known is the experience of making and taking responsibility for choices, whether right or wrong, and seeking to know by this continuing dialogue across the permeable boundary of the self what if anything is worth preserving. It is possible that this quest, given any reasonable degree of responsiveness in the outside world, is what consumers want more than being given what they are told they really want.” (Rothenberg 1962: 282f., italics added)

Put differently, what actually seems to be valuable in a ‘behavioral world’ is making the experience, on an ongoing basis, of trying out new preferences, discarding some of them, and keeping others. This may be referred to as a process of preference learning. It resonates with the notion, advanced by James Buchanan in his *Natural and Artifactual Man* (Buchanan 1999), that human beings face the existential challenge to creatively construct their own identity over the course of their lifetime. By interacting with their environment and especially by engaging in active choice, agents reconstitute themselves again and again. Without this
process, they would be unable to act on reasons. As Buchanan puts it, “the ‘individual’, as described by a snapshot at any given moment, is an artifactual product of choices that have been made in prior periods, both by himself or herself and others” (Buchanan 1999: 287).

Does this approach have any normative bite? It seems that it does. First of all, in the sense of negative policy advice (that prevents harm, rather than fostering the allegedly ‘good’): Consider bias-exploiting nudges, both in their private (commercial) and public (governmental) use. A key risk of nudging is that by systematically discouraging people from engaging in active choice, they risk compromising people’s identity formation, thereby ‘infantilizing’ them (White 2013). The approach also has some normative implications with regard to positive policy advice: by emphasizing the aspect of ‘preference learning’, it favors practical policies that foster people’s ‘opportunity to learn’ rather than circumventing or de-biasing their decision-making. When assessing institutional settings, one should ask: “Do they allow individuals to try out and learn new preferences?” rather than “Do they allow individuals to satisfy their true preferences?” In order to truly inform policy-making, we advocate normative economics to conceptualize a behaviorally robust model of preference formation, rather than holding on to a choice-based understanding of welfare.

We don’t pursue this avenue further at this point. Suffice to say that in light of the recent surge in interest in behavioral policy advice and the shortcomings of purely ‘pragmatic’, ostensibly value-free, approaches to develop such advice, we are confident that the necessary debate on Behavioral Normative Economics will pave the way towards a renewed interdisciplinary dialogue between economics and ethics.

---

31 See also Korsgaard (2009).

32 This argument is consistent with the notion of “boost policies” that is advocated by proponents of the simple heuristics program, see, for one, Gigerenzer (2005) or Grüne-Yanoff and Hertwig (2015).

33 For an in-depth discussion of preference learning, see Schubert (2012, 2015a).
References


