Reconciling Normative and Behavioural Economics: A Critical Survey

Ivan Mitrouchev

Univ Lyon, UJM Saint-Etienne, GATE UMR 5824, F-42023 Saint-Etienne, France
Université de Reims Champagne-Ardenne, REGARDS EA 6292, 51571 Reims Cedex, France

April 2021*

Abstract

The issue of finding a normative approach that is consistent with behavioural economics has been labelled as the 'reconciliation problem' by McQuillin and Sugden (2012). Yet as the authors put it, 'nothing remotely like a consensus has been reached about how the problem is best tackled' (p. 554). This paper is an attempt at suggesting a consensus on how the reconciliation problem can be best tackled. It aims at giving it a unified structure so that some important normative criteria proposed in response to the reconciliation problem — namely *experienced utility*, *true preference*, *choice-basis* and *opportunity* — can be assessed by a simple framework. The framework I propose is based on three requirements that a 'good' normative criterion should satisfy (that I call the *general*, *ethical* and *practical* requirements). After surveying the methodological and theoretical issues of the *experienced utility*, *true preference*, *choice-based* and *opportunity* criteria, I show that none of these normative criteria however satisfy all three requirements. This opens perspectives of further research on debating whether some other alternative normative criteria (*virtue ethics* and *meaning*) could be promising candidates to solve the 'reconciliation problem'.

Keywords. *choice — cognitive biases — normative criterion — preference — behavioural public policy — value judgement*

JEL codes. B41, D63, D90, I31

*A preliminary draft of this paper was presented at the 2019 conference on Public Economic Theory in Strasbourg and at the 2019 Happiness Economics and Interpersonal Relations conference in Rome. I thank the audience for comments. I also thank Jean-Sébastien Gharbi, Francesco Guala, Cyril Hédoin and Robert Sugden for helpful remarks. In particular, I am grateful to Wade Hands for insightful discussions that considerably led me to improve the ideas presented in this paper. All mistakes remain my own.

1 Introduction

Over the past fews decades, behavioural economics (more particularly the heuristics-and-biases program) has documented many preference 'irregularities', which support the view that individual behaviour is, in many cases, not aligned with the norms of rational choice (see Kahneman and Tversky (2000) and DellaVigna (2009) for extensive reviews). These recurrent findings are not only concerning for positive economics, which for around three quarters of a century has been grounded under the assumption of rational choice.[1] They also have an impact on normative economics, which typically considers the satisfaction of 'well-ordered', 'coherent', 'rational' or 'integrated' preferences as what constitutes individual welfare. In particular, if (i) there is recurrent empirical support for the view that individuals do not behave according to the norms of rational choice, and if (ii) rational choice (or the satisfaction of coherent preference) is supposed to define individual welfare, then (iii) it is not clear anymore what aspect of individual behaviour should define what makes individuals better off. This problem of 'reconciling behavioural with normative economics' has been coined as the 'reconciliation problem' by McQuillin and Sugden (2012). The essential question is, how to find a way to do normative economics that is consistent with behavioural economics (or that is consistent with the observation that individuals' preferences are incoherent)?

So far, the reconciliation problem has been interpreted through many perspectives. To name a few, some authors focus on the identity problems of individual welfare (Lecouteux 2015; Dold and Schubert 2018; Mitrouchev and Buonomo 2020), others on various issues of libertarian paternalism (Grüne-Yanoff 2012; Guala and Mittone 2015; Hédoin 2015; Sugden 2017b; Hands 2020), others on quantitative techniques for eliciting individuals' 'true' preferences (Bleichrodt, Pinto, and Wakker 2001; Pinto-Prades and Abellan-Perpiñan 2012), others on questioning the assumption of 'true' preferences (Sugden 2015; Lecouteux 2016; Infante, Lecouteux, and Sugden 2016), others on theoretical models for identifying mistakes (Köszegi and Rabin 2007, 2008; Beshears et al. 2008; Bernheim 2016), others on the relevance of the preference-satisfaction criterion itself (Hausman 2012, 2018; Hédoin 2017), yet others on the ambitious challenge of reconstructing normative economics without the concept of preference (Sugden 2004, 2018). In the same way Bernheim (2016) introduces his 'unified' approach to behavioural welfare economics, one may say that 'reading the literature, one can take the impression that the [reconciliation problem] has become a bit of free for all' (p. 13).[2]

The aim of this survey is to give the reconciliation problem a 'unified' structure so that economists and philosophers who are interested in this topic of research can advance on the important question of how to find a way to do normative economics when individuals have incoherent preferences. Such a contribution is needed for at least two reasons. First, and to the best of my knowledge, there is simply no literature review

[1] The standard axioms of rational choice are *completeness* [$\forall x, y \in X, x \succcurlyeq y$ or $y \succcurlyeq x$], *transitivity* [$\forall x, y, z \in X, x \succcurlyeq y$ and $y \succcurlyeq z \implies x \succcurlyeq z$], *stability* [individuals' preferences are stable over the period for which the external observer observes their choice behaviour], and *context-independency* [preferences should not vary depending on the context of the choice, or in the manner the choice is presented].

[2] In the original quote, Bernheim (2016) refers to behavioural welfare economics, not to the reconciliation problem. But if behavioural welfare economics is one answer (among others) to the enquiry of finding a normative approach that is consistent with behavioural economics, it is only a subset of the reconciliation problem.

on the reconciliation problem.[3] Second, there seems to be a conceptual fuzziness about what sort of consensus the reconciliation problem can reach on how the problem is best tackled. Such a difficulty is well emphasised by McQuillin and Sugden (2012, p. 554), who do not aim at proposing a clear-cut framework of the reconciliation problem but who only draw its main guiding line. Fundamentally, normative economics is based on the use of normative criteria to evaluate individual or social states of affairs, and then to recommend/prescribe public policies based on such evaluation. We may then see the reconciliation problem to be essentially about finding *normative criteria* that are consistent with behavioural economics. The point is if there is a reconciliation problem to be solved, it may be useful — at least from a methodological viewpoint — to provide an answer to the fundamental question of *what a good normative criterion is*.

By 'unified', my aim is to propose a framework on which economists and philosophers could consensually agree on what it takes for a normative criterion to be a 'good' normative criterion. The idea is once we consensually agree on some basic requirements that a normative criterion should satisfy, all we have to do is to assess whether one normative criterion satisfies these basic (or fundamental) requirements. The idea is if all requirements are satisfied, the normative criterion is judged to be 'good'. Otherwise it is not.[4] The issue that there is yet no consensus on how the problem is best tackled can then be formulated as follows. *What basic requirements should a normative criterion satisfy so that it is considered to be a candidate for 'solving' the reconciliation problem?* My suggestion is that these basic requirements can take the following forms. A normative criterion is judged to be 'good' if:

1. **General requirement.** It can be applied to a wide range of choice situations.
2. **Ethical requirement.** It can capture the many different aspects of life that individuals can find valuable.
3. **Practical requirement.** It can measure individuals' states of affairs using a relatively consensual measure of what makes individuals better off (to be discussed below).

By definition, a normative criterion is a rule that tells us whether one outcome is better than another. We have then three particular requirements, which account for three particular problems: 'better when?', 'better to what?' and 'better how?'. Answering the first question implies to have an overall idea of the domain in which a given normative criterion applies. To define the 'normative-relevant' domain is necessary because we need to delimitate the boundaries of the normative relation R, i.e. what it can and cannot evaluate.[5] Answering the second question implies to have an ethical judgement over

[3] McQuillin and Sugden (2012) is of course the main reference but as the authors put it, 'Our intention is to focus on the nature of the reconciliation problem and on the general approaches that might be used in trying to solve it; it is not to write a comprehensive survey article or literature review' (p. 554).

[4] I acknowledge that the major difficulty of such a methodological assessment — that is based on meta-criteria that define what a 'good' normative criterion should be — is obviously the problem of *who* should judge what a 'good' normative criterion should be. The difficulty of finding a consensus among economists and philosophers not only presupposes that most of them agree with these meta-criteria (that I call 'requirements'), but also that these meta-criteria are postulated *ex-ante* (and not *ex-post*). I leave these important considerations for the discussion section.

[5] I deliberately do not use the usual vocabulary of 'welfare-relevant domain' nor 'welfare-relation', because a comparison is not necessarily based on *well-being* (or welfare, or individual utility). Otherwise, normative economics would be restricted to a narrow range of application.

the normative-relation 'better than'. To define a normative relation R, outcomes x and y, and to say that $x \; R \; y$ means that 'x is better than y' is mathematically purposeful, but meaningless if we do not define the content of this normative relation. Answering the third question implies to have a measure of this ethical content that allows to practically evaluate individuals' states of affairs, without which no evaluation would be possible. In sum, the general requirement is the the ability of a normative criterion to apply to a wide range of choice situations (the *scope* of the normative-relevant domain); the ethical requirement is the ability of a normative criterion to 'cut up the world', i.e. to judge what situation is considered to be better than another regarding individuals' interests (the *content* of the normative-relevant domain); the practical requirement is the ability of a normative criterion to measure individuals' states of affairs (the *measurability* of the normative-relation). My suggestion is that these requirements could provide solid grounds for paving the way towards a unified account of the reconciliation problem, and eventually for advancing on the way to 'solve' the reconciliation problem.

The rest of the paper is organised as follows. Section 2 presents the reconciliation problem. Section 3 critically surveys the main normative criteria offered in the literature, which can be divided in two categories: individual maximisation-based criteria (*experienced utility*, *true preference*, *choice-basis*) and freedom-based criterion (*opportunity*). These normative criteria can be seen as responses to the reconciliation problem because they offer solutions to the problem of evaluating individuals' states of affairs when individuals' preferences are incoherent. Section 4 then assesses the *experienced utility*, *true preference*, *choice-based* and *opportunity* criteria with respect to the three requirements I here propose. As we shall see, if these requirements are accepted then none of these normative criteria satisfies them all. This leads me to suggest in Section 5 avenues of future research on debating about the status of these meta-criteria (that I call 'requirements'), and on discussing other less known normative criteria that could perhaps better satisfy these three requirements. Section 6 concludes.

2 The Reconciliation Problem

Standard normative economics — whose archetype can be labelled as standard welfare economics — conventionally uses preference-satisfaction as a reliable indicator of well-being and assumes that individuals have coherent preferences. The satisfaction of individuals' coherent preferences is then considered to be the proper normative criterion for evaluating individuals' states of affairs. However, when individuals' preferences are incoherent due to numerous reasons (non-Bayesian updating, framing, non-exponential discounting, intransitivity, *status quo* bias, violation of dominance, etc.), economists do not have a stable normative criterion so that they can evaluate individuals' states of affairs. That is to say, if the assumption of coherent preference breaks down then the standard normative criterion of preference-satisfaction is no more relevant to indicate what makes individuals better off.[6] As McQuillin and Sugden (2012) put it, the con-

[6] By 'no more relevant', I refer to situations where individuals are unlikely to have coherent preferences. There is of course the argument that if decision tasks are well-specified and are repeated a sufficient number of times — the 'discovered preferences' hypothesis (Plott 1996) — the repetition of the game in a market environment enables individuals to get more experience so that they can *in fine* 'discover' their coherent preferences. What can nonetheless be said is that the documentation of incoherent preferences for decisions which are not sufficiently repeated — and actually also for some which are sufficiently re-

cept of preference-satisfaction in standard welfare economics used to be relatively unproblematic as long as we assumed that individuals had coherent preferences. In fact, preference-satisfaction could be justified under the following three interpretations.

- *Happiness interpretation*. Individuals seek to maximise their happiness.
- *Well-being interpretation*. Individuals act according to their well-being (that is not necessarily equal to their happiness).
- *Freedom interpretation*. Individuals act with respect to the freedom of making their own decisions (in line with the neoclassical consumer sovereignty principle).

All these interpretations attribute to individuals a coherent way of acting towards their own interests. In this case, economists did not have to bother about having a normative criterion at hand apart from preference-satisfaction. Indeed, albeit these interpretations provided different reasons that preference-satisfaction mattered, substantive ethical questions about *why* preference-satisfaction mattered did not to be asked under the assumption of coherent preference. As McQuillin and Sugden (2012) put it,

> 'economists could use a common theoretical system in which preference-satisfaction was the normative standard, while disagreeing about *why* preference-satisfaction mattered.' (p. 555 — their emphasis)

But once we seriously take into account that preferences are incoherent, interpretations about why preference-satisfaction matters — either in terms of *happiness*, *well-being* or *freedom* — suggest very different answers to the reconciliation problem. Under the empirical evidence that preferences are incoherent, it seems that we cannot seriously hold that individuals seek to maximise their happiness because we may find out that individuals do not choose according to what makes them happier. Instead, we may have to propose alternative measures of happiness that are not related to observed choice but to *hedonic experience*. The well-being interpretation faces the same faith. If preferences are incoherent, well-being cannot be grounded on the standard framework of observed preferences anymore since the large amount of cognitive biases documented in the literature suggests that individuals may not always act according to their own interests. Instead, we may have to attribute some preferences 'special' properties so that they can provide a better measure of one's well-being — typically the property that such preferences are *undistorted by cognitive biases*. The freedom interpretation perhaps provides the most straightforward answer to the reconciliation problem. If preferences are incoherent, we may simply have to disentangle the idea that it is good to satisfy individuals' preferences because it is *their* preferences (the consumer sovereignty principle) and the preference-satisfaction concept. In other words, rather than assuming that the consumer sovereignty principle depends on coherent preference, we may instead focus on the institutional process that allows individuals to *enhance their opportunity to choose from*, disregarding whether their preferences are coherent or not.

Note that the problems associated with preference-satisfaction are far from being new. Eminent contributors to welfare economics and social choice such as Harsanyi (1977) and Sen (1991) already aimed at giving preference-satisfaction a substantial

peated *and* made by market experts (Dhami 2016, p. 1449) — is significant enough to think about how normative economics should handle incoherent preferences.

content without waiting any reconciliation problem to be recognised as such. Also, there have been long standing concerns about individual preference-satisfaction: whether it makes any sense as a definition of well-being and its practical usefulness.[7] However, the accumulation of empirical evidence that individuals' preferences are incoherent is now large enough to recognise the reconciliation problem as a topic of its own. The point is, if normative economics needs to be consistent with behavioural economics then normative economics should ground normative criteria on something other than observed choice. These normative criteria can be grounded on the three ethical loci of *happiness*, *well-being* and *freedom*, but nothing forbids to ground normative criteria on yet other ethical loci that are unfamiliar with the tradition of normative economics (to be discussed in Section 5). The aim of the next section is to review some of the important methodological and theoretical issues associated with the main normative criteria to solve the reconciliation problem. These normative criteria are either based on individual utility maximisation — and thus take either the happiness or well-being interpretation of preference-satisfaction (*experienced utility*, *true preference* and *choice-basis*) — or break out with the concept of preference-satisfaction by taking the freedom interpretation (*opportunity*).

3 Solving the Reconciliation Problem: A Critical Review

3.1 Individual Utility Maximisation-Based Criteria

3.1.1 Experienced Utility

The *experienced utility* criterion takes the happiness interpretation of the reconciliation problem. To contextualise things from a history-of-economic-thought perspective, it — strictly speaking — aims at measuring individuals' hedonic states in the line of Bentham's (1780 [2007]) meaning of utility as pleasure/pain calculus and of Edgeworth's (1881) concretisation of a hedonimeter.[8] Knowing that the 'grand hedonistic tradition' dominated early neoclassical economics for its (roughly) first fifty years, it is worth noting that experienced utility — explicitly presented as a 'back-to-Bentham' approach — is just a special case of such hedonistic tradition. The conceptual appeal of measuring individuals' hedonic states is based on the theoretical discrepancy between what individuals do (what the authors refer to as *decision utility*) and what they experience (what the authors refer to as *experienced utility*). Since what individuals do is subject to many cognitive biases, the idea is only to take what they experience in terms of pleasure and pain as the benchmark for evaluating their states of affairs. The ethical premise of the experienced utility criterion can thus be formulated as follows.

Experienced utility (ethical premise). *It is good to maximise individuals' experiences of pleasure (or to minimise individuals' experiences of pain).*

[7] See e.g. Hausman, McPherson, and Satz (2016) for how problematic individual preference-satisfaction is as a standard for well-being, without seriously addressing the impact of behavioural economics.

[8] The literature on measuring experienced utility includes Kahneman and Snell (1990, 1992), Kahneman and Varey (1991), Varey and Kahneman (1992), Kahneman et al. (1993), Fredrickson and Kahneman (1993), Kahneman (1994, 1999, 2000, 2011 [Part V]), Redelmeier and Kahneman (1996), Kahneman, Wakker, and Sarin (1997), Schreiber and Kahneman (2000), Redelmeier, Katz, and Kahneman (2003), Kahneman et al. (2004), Kahneman and Sugden (2005), Kahneman and Krueger (2006), Kahneman and Thaler (2006) and Dolan and Kahneman (2008).

The methodological and theoretical problems of the experienced utility criterion are various, but I shall restrict myself to the ones that are perhaps the most concerning (see Mitrouchev (2020) for a detailed assessment).

First, it is often argued that hedonism, when formulated into the maximisation of experienced utility, is too narrow a criterion to capture what makes the good life. This point is well acknowledged by tenants of the experienced utility criterion, who argue that hedonic experience is a component of the good but that the good life is certainly not empty of hedonic evaluation.[9] This is however already problematic for the ethical and general requirements because (i) there is a wide range of dimensions of the good life that is neglected with this normative criterion and (ii) one may wonder whether it is the goal of public policy to promote pleasurable experiences and not indirect measures of happiness, such as access to public transport, green spaces, good air quality, etc. — and let individuals free to pursue whatever they want.

Second, experienced utility resurrects an ancient evil of standard welfare economics: interpersonal comparisons of utilities. The theory of experienced utility measurement is constructed of several axioms (Kahneman, Wakker, and Sarin 1997; Kahneman 2000). One axiom strictly assumes ordinal comparisons between individuals' utilities of different outcomes (e.g. one individual experiences the taste of an exotic fruit, the other experiences a guiding tour in Louvre Museum), while also assuming cardinal comparisons between individuals' utilities of the same outcome (e.g. two individuals experience the taste of the same ice-cream). This may be concerning for those who are reluctant to hold that there is something 'scientifically relevant' to interpersonal comparisons of utilities, especially when cardinality is *stricto sensu* assumed. Of course, debating on whether cardinality is scientifically relevant for both positive and normative economics would lead us to a well-known and long strand in the history of economic thought. The point is albeit it could make sense to consider experiences of *pain* to be commensurable, it is far from absurd to consider experiences of *pleasure* to belong to another system of perception that works differently. This is due to some psychological phenomena such as the 'adaptation effect': individuals adapt to their circumstances to the point that they do not perceive an increase of pleasure similarly than another individual who is differently endowed (e.g. a rich individual compared to a poor individual).[10]

Third, experienced utility measurement has been given reconsideration by Kahneman himself in a recent interview given to *Hareetz* newspaper (Kahneman 2018). The theoretical construction of experienced utility measurement is grounded on *moment utility*: what is experienced here and now. The main point made by Kahneman (1999) is that 'policies that improve the frequencies of good experiences and reduce the incidences of bad ones should be pursued even if people do not describe themselves as happier or more satisfied' (p. 15). In other words, the author argued that only the maximisation of moment utility is normatively relevant, *even if individuals actually have a more pleasant memory of an experience with lower moment utilities*. However, nothing really says why

[9]See Varey and Kahneman (1992, p. 169), Kahneman (1994, p. 21), Kahneman, Wakker and Sarin (1997, p. 377) and Kahneman and Sugden (2005, p. 176), who make that point explicit.

[10]Kahneman (1999) however provides an extensive psychological defence on how to deal with this issue, not without acknowledging the methodological difficulties.

moment utility should be given more importance than *remembered utility* (the global retrospective evaluation of a past experience). In fact, Kahneman recently reconsidered his position, claiming that hedonic measurement may not be what matters to individuals' objective happiness. As the author puts it:

> 'People don't want to be happy the way I've defined the term — what I experience here and now. In my view, it's much more important for them to be satisfied, to experience life satisfaction, from the perspective of "what I remember", of the story they tell about their lives. I furthered the development of tools for understanding and advancing an asset that I think is important but most people aren't interested in.' (Kahneman (2018) — interviewed)

This statement of Kahneman after (approximatively) twenty years of research in experienced utility measurement may yield to a considerable change in our understanding of how objective happiness should be measured. Perhaps it should not be measured in terms of internal states of mind (immediate pleasure) but rather in terms of what individuals remember of their experiences (which is a more durable state of mind than immediate pleasure), or even of consideration about what makes the good life that is not directly related to pleasure (e.g. access to public transport, green spaces, quality of the air, etc.). The latter interpretation of happiness refers to an objective conception of goodness, where pleasure is only a by-product of happiness but does not *constitute* happiness.

3.1.2 True Preference

Unlike the experienced utility criterion, the true preference criterion does not take such a tangible account of what matters to individuals (i.e. happiness defined in terms of a pain/pleasure calculus) but assumes a general psychological state in which individuals have the ability to meet their actual intentions/interests that are represented by their true/latent/laundered preferences. From the social planner's viewpoint, these preferences constitute individuals' 'normative' preferences (i.e. what they *should* prefer).[11] *True* preferences (by opposition to other kinds of preferences that would be 'false', or 'mistaken') are defined as preferences that an individual would have had, had she not been disturbed by rational foibles/biases/errors/mistakes/anomalies/cognitive disturbances. The representation of actual choice as a combination of true preferences and errors allows the social planner to only take true preferences as normatively relevant, i.e. as what constitutes individuals' well-being. The social planner's goal is to identify these errors and then to reconstruct/recover individuals' true preferences through various social mechanisms.[12] One main advantage of this normative criterion is that it does

[11] The literature that uses true preference as a standard of well-being includes Bleichrodt, Pinto, and Wakker (2001), Madrian and Shea (2001), Benartzi and Thaler (2002), Camerer et al. (2003), Thaler and Sunstein (2003, 2009), Thaler and Benartzi (2004), Beshears et al. (2008), Loewenstein and Haisley (2008), Dalton and Ghosal (2011), Rubinstein and Salant (2012), Pinto-Prades and Abellan-Perpiñan (2012), Halpern (2015), Thaler (2018) and Sunstein (2019).

[12] The most eminent approach is certainly the *nudge* literature (Thaler and Sunstein 2009; Halpern 2015). Note that before a large amount of behavioural economists had a great interest in the true preference criterion, some authors already gave considerable support for it. The concept of true preference follows the one of Harsanyi (1977, pp. 29-30) in his defence of utilitarianism. Fine (1995) aimed at distinguishing the two concepts of true preference and actual choice from a social choice perspective. From a philosophical perspective, some authors already defended the satisfaction of self-interested 'informed', 'rational', or 'laundered' preferences as what constitutes goodness (Gauthier 1986, Ch. 2; Arneson 1990; Goodin 1992). See also Railton (2003): 'an individual's good consists in what he would want himself to want, or to pursue ... free from cognitive error or lapses of instrumental rationality' (p. 54).

not require to interpret well-being as narrowly as the experienced utility criterion. In this manner, it may capture different aspects of life that individuals can find valuable. Moreover, it may also be more generalisable to various choice situations. With this approach it is (presumably) up to individuals to define what their own well-being is. The ethical premise of the true preference criterion can then be expressed as follows.

True preference (ethical premise). *It is good to satisfy individuals' preferences that are undistorted by cognitive biases*.

Several issues are however associated with this normative criterion. First, contrary to the experienced utility criterion that is psychologically well grounded — no doubt pain and pleasure are real psychophysical phenomena that can somehow be measured — the true preference criterion shares nothing of this sort. To the question of whether there is empirical evidence for the existence of true preference, we can straightforwardly say that no empirical study has so far supported this claim, nor actually the contrary. The fact that true preference lacks of psychological explanation is the main concern of Infante, Lecouteux, and Sugden (2016), who point out two problematic principles of this normative criterion: (i) even in possession of full cognitive capacities the latent process of producing true preferences is left unexplained, and (ii) decision theory has no competence to legitimise a single correct way of framing a choice problem, which is accessible to any individual (even if 'super-rational'). Perhaps the most important issue of the true preference criterion is its inability to provide a convincing account that individuals would be better off if they satisfy their coherent preferences (Berg and Gigerenzer 2010). In fact, it appears that true preferences seem to be nothing else than *coherent* preferences. The fact that the heuristics-and-biases *positive* program is presented as a radical departure from the *homo-oeconomicus* abstraction but that its *normative* program is (perhaps paradoxically) based on it, is commonly recognised in the critical literature.[13]

Second, if the assumption of true preference lacks of psychological explanation, tenants of the true preference criterion may need alternative ways for justifying its application. One way for the true preference criterion to be operationalisable is to provide the social planner with some meta-criteria on what is judged to be a better outcome over another. This is to avoid the problem that the social planner cannot *know* what individuals' true preferences are.[14] Several meta-criteria have been proposed by tenants of the true preference criterion, but none of them seem to satisfy the general requirement because they restrict the scope of the true preference criterion to a narrow range of application. These meta-criteria are the following.

- *Dominance.* When one alternative strictly dominates another either in terms of outcome or risk, it may be assumed that the former is better than the latter. For example, we may assume that individuals' true preferences are to save the maximum amount they can (e.g. they prefer more money to less when they will be retired). Based on this assumption, the social planner could set the maximum amount as the default option of the 401(k) plan. This meta-criterion is proposed by Loewenstein

[13]See Berg (2003, p. 431), Berg and Gigerenzer (2010, pp. 147-148), Hands (2014, p. 398), Whitman and Rizzo (2015), Lecouteux (2016) and Dold and Schubert (2018).

[14]Rizzo and Whitman (2009) call this problem the 'knowledge' problem and Rebonato (2012) the 'interpersonal intelligibility of preferences' problem.

and Haisley (2008). The issue is that dominance can only apply to some circumstances, where more can unambiguously be compared to less (typically monetary outcomes). Furthermore, the 'more is better' maxim may not necessarily be consensual among individuals. For example, one may not necessarily prefer the travel trip bundle $\{France, Italy, England\}$ to $\{France, Italy\}$, simply because she may not like to visit England. The disliked alternative added to the bundle (here $England$) may play out negatively in the individual's personal evaluation.

- *Evidential view (or folk beliefs)*. This meta-criterion holds the idea that the choice architecture (or framing) is legitimised when there are 'good' reasons to believe that the behaviour being encouraged will actually improve the well-being of individuals being influenced by the social planner. For example, under the assumption that eating healthy, not smoking and saving more is better, the choice architecture should be framed in a way that will encourage individuals to eat healthy, not smoke and save more. Tenants of the true preference criterion who support this meta-criterion in fact mean something closely related to Hausman's (2012) 'evidential view'. The 'evidential view' states that preference-satisfaction does not constitute well-being but provides reliable information about well-being. Instead of having an ethical theory at hand, the idea of Hausman is that folk beliefs about what constitutes goodness are enough to make sense of what makes individuals better off. The platitudinous character of the 'evidential view' is fully recognised by Hausman (2012), who argues that 'platitudes concerning what is good for people still have content ... economists know enough about the things that make lives good or bad that they can make sense of the evidential view of the relationship between preference-satisfaction and welfare' (pp. 92-93). It is however disturbing to see that the author also holds elsewhere that 'economists who believe that they promote well-being by satisfying purified preferences need to know what people's purified preferences are, not what they should be' (Hausman 2016, p. 28). The issue is that folk beliefs only allow to say what individuals' preferences *should be*, not what they *actually are*. Strictly speaking, characterising such a meta-criterion as 'evidential' seems misleading: what kind of 'evidence' folk beliefs provide about what makes individuals better off?

- *Self-officiating (or 'as judged by themselves')*. We are then left with what individuals would express what they judge to be their own good. This meta-criterion is given the name of 'self-officiating' by Loewenstein and Haisley (2008) and 'as judged by themselves' by Thaler and Sunstein (2009). It states that on the condition that individuals clearly self-express their willingness to lose weight, stop smoking, stop procrastinating, etc. then the true preference criterion applies.[15] For example, if overweighted individuals consistently state that they would be better off if they were slim, and if they deliberately state that a paternalistic policy would make them better off, then such a policy would be ethically justified (Loewenstein and Haisley 2008). We can see two problems with this meta-criterion: a philosophical and a practical one. From a philosophical viewpoint, if preference reversals are observed in individuals' statements (e.g. individuals claim to prefer A to be B today and B to A tomorrow) then one may fairly question which of the many individuals' preferences over time has/have moral authority over the other(s) (Mitrouchev and

[15] See Reisch and Sunstein (2016), Reisch, Sunstein, and Gwozdz (2017) and Sunstein, Reisch, and Kaiser (2019) for empirical surveys about Europeans' acceptance of *nudges*.

Buonomo 2020). From a practical viewpoint, one may also argue that economists or social planners specifically want to have a normative criterion at hand when these *ex-post* feedbacks are unavailable (Bleichrodt, Pinto, and Wakker 2001).

- *Clearly negative outcomes.* This meta-criterion states that when everyone would agree under 'common sense' that one outcome is clearly at the cost of individuals' interest, it appears relatively unambiguous that the true preference criterion applies (Loewenstein and Haisley 2008). In other words, this meta-criterion is basically a strong version of the 'evidential view'. Addiction, bankruptcy or paying the exact same product more expansively that it could have been afforded (ethical considerations such as fair trade or environmental protection left apart) are examples of clearly negative outcomes. Among all the meta-criteria here listed, this one has perhaps the most intuitive appeal: some things are just bad for everyone. Perhaps a tiny fraction of people would argue the contrary, but an overwhelming majority seems enough to make the true preference criterion operationalisable under the 'clearly negative outcomes' meta-criterion.

Considering these restrictions on the applicability of the true preference criterion, it follows that the true preference criterion only makes sense in situations where distortions from rationality uncontroversially make individuals worse off. To take yet another example, consider a case where individuals would have to exploit their cognitive abilities more carefully (e.g. studying and comparing prices) if some markets such as water supply, food or phone contracts were not regulated (Heidhues, Johnen, and Köszegi 2020). This can eventually lead individuals to choose a complex tariff that does not minimise their costs. But it would be relatively uncontroversial to state that for a given bundle of alternatives (such as a phone contracts), individuals who have the choice between complex tariffs simply want to choose the one that minimises their costs. The point is that in some specific cases, a good choice is indeed (and uncontroversially) a choice undistorted by cognitive biases. This approach is given the name of 'preference regularisation' instead of 'preference purification' by Infante, Lecouteux and Sugden (2016, pp. 18-21). The authors argue that the true preference criterion only makes sense in such uncontroversial cases.

Third, the oxymoron character of libertarian paternalism (Thaler and Sunstein 2003, 2009) — which is the dominant approach that takes true preference as its normative criterion — is subject to some philosophical problems. From a philosophical viewpoint, when the social planner exploits individuals' cognitive biases to help them taking the best decision, it is difficult not to violate individuals' liberal principles. This is true even if individuals explicitly agree with the paternalistic character of the policy intervention. In other words, libertarian paternalism inevitably requires to make trade-offs between well-being and liberty/freedom/autonomy. But one can hardly increase the former while leaving the latter unchanged.[16]

[16] For the difficulty of combining paternalistic with liberty/freedom/autonomy values, see Mitchell (2005), Rizzo and Whitman (2009, 2019), Welch and Hausman (2010), Grüne-Yanoff (2012), Rebonato (2012, 2014), Hédoin (2015, 2017), Sugden (2017b) and Scoccia (2019).

3.1.3 Choice-Basis

Like the true preference criterion, the *choice-based* criterion also takes the well-being interpretation of preference satisfaction, although it could also be associated with the freedom interpretation. This normative criterion can be seen as a subtle version of the true preference criterion, since it suggests a compromise between the problem that actual choice diverges from well-being and the possibility to nonetheless keep *choice* as a satisfactory indicator of well-being (thus 'rescuing' somehow the consumer sovereignty principle). The choice-based criterion takes a step farther in not defining what makes individuals better off because it only considers a minimal psychological state of observation, attention, memory, forecasting or learning processes for normative assessments (thus leaving any ambiguity of the individuals' reasons for choice apart).[17] The conceptual appeal of this normative criterion is shared by some economists who are reluctant to assess individuals' mental states for either epistemic or practical reasons.[18] It is indeed worth to emphasise that economists usually make of *choice*, and not subjective well-being reports, their privileged data. In this manner, they conform to the liberal tradition of standard welfare economics by making choice (or observed preference) the main normative criterion for well-being. The ethical premise of the choice-based criterion can be formulated as follows.

Choice-basis (ethical premise). *It is good that individuals choose undistorted by cognitive biases.*

As we can see, the ethical premise of the choice-based criterion is almost identical with the ethical premise of the true preference criterion. The subtlety is that the privileged data is here not *preference* but *choice*. That is, the social planner is not required to have individuals' expression of preference but only to identity cognitive anomalies with the help of theoretical models that formally define what a mistake is (Köszegi and Rabin 2007, 2008; Bernheim 2016). The goal of the social planner is then only to take observed choice undistorted by cognitive biases as the normative-relevant data. The main issues associated with this normative criterion are nonetheless the following.

First, if the choice-based criterion is grounded on the same assumption than the true preference criterion (i.e. that what makes individuals better off is some psychological states that are free from cognitive biases), then how does it fundamentally differ with the true preference criterion? The simple answer is that (fundamentally), there seems to be no difference between these two normative criteria. The choice-based criterion therefore meets the same critique of the true preference criterion: it only accounts for situations in which distortions from rationality 'obviously' make individuals worse off.

[17] The literature using the choice-based criterion as a standard of well-being includes Bernheim and Rangel (2004), Köszegi and Rabin (2007, 2008), Bernheim and Rangel (2007, 2008, 2009), Salant and Rubinstein (2008), Loewenstein and Ubel (2008), Bernheim (2009), Dalton and Ghosal (2012), Rubinstein and Salant (2012), Masatlioglu, Nakajima, and Ozbay (2012), Manzini and Mariotti (2014) and Bernheim (2016).

[18] See Bernheim and Rangel (2008, p. 156), Manzini and Mariotti (2014, pp. 343-344) and Bernheim (2016, pp. 24-25), who advance the classical 'ordinalist' argument, according to which welfare economists should evaluate individuals' states of affairs based on individuals' own conception of goodness (not happiness nor true preference). They also argue that choice is a far less obscure and more available data than anything else.

Second, although tenants of the choice-based criterion are reluctant to assess individuals' states of affairs by measuring individuals' level of happiness, they still make room for mental states by giving it an 'auxiliary role' (Manzini and Mariotti 2014, p. 344). But then the choice-based criterion seems to encounter a disturbing paradox that is well emphasised by Dhami (2016):

> 'Choice-based models must address the issue of choices that depart from those expected under the rational benchmark. In a leading model, one deals with this issue by trimming-away the anomalous choices. However, such trimming-away necessitates the use of either non-choice data, or the invocation of a welfare criteria for trimming the choices, which is what one is trying to construct in the first place.' (p. 1577)

The issue is if the choice-based criterion ultimately depends on either experienced utility or true preference, what is (fundamentally speaking) the added value of this normative criterion?

Third, perhaps the most concerning issue of the choice-based criterion is that it is (presumably) not ethically grounded at all. Unlike the other normative criteria based on individual utility maximisation here reviewed (*experienced utility* and *true preference*), only tenants of the choice-based criterion refuse to say something on the ethical content of the normative-relevant domain. This position is explicitly defended by Bernheim (2016), who holds the usual 'ethically neutral' stance of standard welfare economics. In his words, 'The conventional economic framework seeks to assess well-being without factoring in ... moral considerations, concerning which economists have no special expertise. I follow that tradition' (p. 18). But since normative criteria are, by definition, rules that tell us whether one outcome is better than another, there is merely no way of avoiding ethical judgements about what makes one outcome actually better than another. The question is, how can a normative criterion be 'normative' at all if it does not presuppose what makes one outcome better than another? Yet claiming that individuals are better off when they are undistorted by cognitive biases already seems to be an ethical judgement regarding what constitutes goodness.

3.2 Freedom-Based Criterion

3.2.1 Opportunity

The *opportunity* criterion breaks with rational choice as the normative benchmark, which is a common point shared by the *true preference* and *choice-based* criteria. Recall that the true preference and choice-based criteria lean on a separation between rational reasoning and cognitive biases. By emphasising that incoherent preferences are not incompatible with normative analysis, Sugden (2004, 2018) proposes a normative criterion of *opportunity*, according to which more opportunity for individuals is better than less, *independently of what their preferences are*.[19] The aim of the author is to maintain the liberal tradition of economics against libertarian paternalism, whose purpose is to combine liberal and paternalistic principles (but perhaps unsuccessfully, as previously mentioned). Sugden's (2018) two central criticisms is that there is no reason to assume that true preferences exist beneath the psychology of actual mental processing, and that the social planner's viewpoint is irrelevant because *individuals* (not the social planner) are the

[19] The literature on the opportunity criterion includes Sugden (2003, 2004, 2007, 2008, 2010, 2017a, 2018).

addressees of public policy. The author ambitions to replace what he calls the process of ‘preference purification’ with the concept of ‘opportunity for choice’. While the former aims at reconstructing/recovering individuals’ true preferences, the latter focuses on enhancing individual freedom to choose. In that matter, Sugden’s approach takes the freedom interpretation of preference satisfaction. The benefits of the opportunity criterion are twofold: (i) it avoids the problematic aspects of the true preference and choice-based criteria of determining what a decision ‘free from cognitive biases’ is, and (ii) it avoids saying something about what constitutes goodness by instead leaving individuals be the best judge of their own good. The ethical premise of the opportunity criterion can then be formulated as follows.

Opportunity (ethical premise). *It is good to promote individuals with more opportunities to choose rather than less.*

Like the rest of the normative criteria previously reviewed, the opportunity criterion is however not unproblematic from both methodological and theoretical perspectives.[20] First, the opportunity criterion forbids to make comparisons between sets that are not nested. To give a simple illustration, consider the opportunity set $O_1 = \{x, y, z\}$ compared to the opportunity set $O_2 = \{x, y\}$. Here O_1 dominates O_2 under the opportunity criterion because O_1 contains all the alternatives in O_2 (that is, x and y) *plus* an alternative that is unavailable in O_2 (that is, z). But what if we have an alternative in one opportunity set that is not contained in another one, e.g. $O'_1 = \{x, y, z\}$ and $O'_2 = \{w, x\}$? Because Sugden (2018) does not suggest that the *nature* of some alternative may provide more opportunities than other ones, the opportunity criterion is silent about evaluating opportunity sets that are not nested.[21] The same issue applies for any other combination where one alternative is not contained in another opportunity set. Consider for example a case where $O''_1 = \{r, s, t, u, v, x, y, z\}$ and $O''_2 = \{w\}$. In this case, we can still not say anything on whether it is O''_1 or O''_2 that provides more opportunity, even if the cardinal of alternatives in O''_1 is by far larger than the singleton in O''_2. This constitutes a challenge for the general requirement, as there might be many situations where non-nested sets need to be compared.

Second, there are the psychological phenomena of *choice overload* and *self-constraint*, which may challenge the ethical premise that more choice (or opportunity) is always better than less.

- *Choice overload*. Choice overload is defined as the feeling of being worse off by having too many alternatives to choose from.[22] Schwartz (2004 [2016]) identifies the following negative feelings associated with it:
  - *Paralysis (or inefficiency)*. More alternatives create paralysis (i.e. it is difficult to choose something at all). A related psychological phenomenon is emphasised by Benartzi and Thaler (2002), who show that more opportunities lead to more complexity and then to an inefficiency in picking the best outcome.

[20] What follows applies to Sugden’s *individual* opportunity criterion, not to his *interactive* opportunity criterion. See Sugden (2018) for the different concepts that the opportunity criterion can take.

[21] This point actually refers to a complex debate in social choice on how to measure opportunity, and whether opportunity is measurable at all. I come back to this point below.

[22] The empirical literature includes Iyengar and Lepper (2000), Hutchinson (2005), Botti and Iyengar (2006) and Scheibehenne (2008).

- *Decreasing of satisfaction*. Even if individuals are not paralysed, they may end up being less satisfied than with fewer options. The potential reasons are the following.
  * *Regret and anticipation regret*. An individual who faces too many alternatives would easily imagine what it would have been if she had chosen another alternative. This tends to increase the risk of regretting the chosen alternative.
  * *Opportunity cost*. This refers to the previous reason formulated in economical terms. If the opportunity set is large, it is easy to think about missing an opportunity, thus making the individual less satisfied with the chosen alternative.
  * *Escalation of expectations*. The more choice the individual has, the more demanding she may become. In other words, her expectations may increase with the increasing of available alternatives. This eventually makes her less satisfied than she would have been if she had the choice between fewer alternatives.
  * *Self-blame*. The opportunity criterion is grounded on the consumer sovereignty principle, according to which not only individuals are the best judge of their own well-being but are also fully *responsible* for their own choice (Sugden 2004, p. 1018). Consequently, it becomes easier to blame oneself for not having made the 'right' choice.

Albeit choice overload is documented in many experimental studies, the meta-analysis of Scheibehenne, Greifeneder, and Todd (2010) concludes that such a psychological phenomenon is however not clearly significant. According to the authors, it is important to study all the explanatory variables that may either facilitate choice overload or not. Yet another meta-analysis of Chernev, Böckenholt, and Goodman (2015) identifies four factors that do facilitate choice overload: *choice set complexity*, *decision task difficulty*, *preference uncertainty* and *decision goal*. In a nutshell, unlike the empirical literature on incoherent preference the empirical literature on choice overload provides far less homogeneous conclusions to safely advance that choice overload can be considered as a stylised fact. Nonetheless, it seems not absurd to take the large amount of empirical evidence that gives substantial support for choice overload as an importent aspect of economic reality. One may also argue that the ethical premise 'more is better' highly depends on the *nature* of the alternatives. As Schwartz (2004 [2016], pp. 24-25) puts it, some alternatives are perhaps worth being available in large varieties (e.g. food at the supermarket), while other alternatives may not (e.g. public utilities, education or health insurances). There is indeed no *a priori* reason to assume that all the available alternatives in the economy are not perceived differently among individuals (i.e. either 'less-opportunity wanted' or 'more-opportunity wanted'). In other words, there may be some alternatives that individuals would like to have more opportunity to choose from, but not other.

- *Self-constraint*. This psychological phenomenon is characterised as the explicit want to have less alternatives rather than more.[23] Unlike choice overload, self-

[23]The philosophical literature includes Elster (1979 [1998], 1983 [2016], 2000). See also Thaler (1980), who discusses situations where individuals voluntarily restrict their choices, deliberately not

constraint is something that is explicitly wanted by the individual. Therefore, perhaps it constitutes a bigger challenge to the opportunity criterion, which (again) gives fundamental importance to individual responsibility (i.e. being the master of one's own choice). To illustrate how self-constraint may challenge the opportunity criterion, consider the following case where an individual has two possible consumption alternatives *fruit* and *cake* that she can consume in periods 1 and 2 (Sugden 2018, p 150). The individual can choose between a fruit and a cake in both periods, so her opportunity set is defined as $O = \{\{fruit, cake\}, \{fruit, cake\}\}$. Now assume that the same individual would like to constrain her opportunity set only to *fruit* in period 2 (for some reasons that are not of the concern of the social planner nor anyone else). She can choose between a fruit and a cake in period 1 but only a fruit in period 2. Hence, her opportunity set is defined as $O' = \{\{fruit, cake\}, \{fruit\}\}$. According to Sugden's (2018) individual opportunity criterion, 'any expansion of a person's opportunity set promotes her interests' (p. 99). Therefore, O dominates O'. However, if we ground normative assessment on the consumer sovereignty principle, according to which we must give fundamental importance to the individual's choice because it is *her* choice, then we must respect her will to restrict her freedom to choose and therefore rank her opportunity sets in a way that O' dominates O. Paradoxically, the opportunity criterion suffers from a disturbing theoretical problem: it does not account for the interests of individuals who want to constrain their own alternatives without violating its principle of providing individuals with more choices rather than less. To put it simply, it does not account for individuals who would like to have *the choice of not having the choice*.[24] This limitation of the opportunity criterion is well recognised by Sugden (2018). In his words,

> 'How far a regime of voluntary transactions should be regulated so as to support individuals in imposing constraints on themselves is a deep problem that generations of economists have struggled with. I can only say that my analysis, as I have so far developed it, abstracts from this problem.' (p. 151)

To this objection, we can argue (like Sugden does) that very little economic activities are concerned with self-constraint, so that self-constraint may not constitute a big challenge for the general requirement. But we may retort that individuals do not necessarily need to show an explicit want for self-constraint *before* their decision. Instead, the need for self-constraint may arise once individuals notice they would have been better off with less choice rather than more. In other words, it may be an *ex-post* rather than an *ex-ante* individual evaluation: individuals may simply not be aware of the value of self-constraint *before* figuring out that more choice can make them worse off.

Taking the two psychological phenomena of *choice overload* and *self-constraint* together, one problem of the opportunity criterion seems to be that it is merely *a-psychological*.

choosing so as to avoid psychic costs that the choices might induce.

[24]One may object that as long as the individual has the choice between O and O', my criticism does not apply because the individual is still free to choose whatever she wants. But if we take individual psychology seriously, we should account for the individual's possible conflicting preferences over time. That is, we should account for the fact that she may like to avoid being tempted by the cake, so having the choice between O and O' would not help her because her 'morally responsible self' would only like to have the choice of O'.

That is to say, it avoids giving any normative relevance to individuals' psychological processes when they choose. Yet as initially formulated by McQuillin and Sugden (2012), the reconciliation problem is to find normative criteria that are *consistent* with individual psychology. If it is so, the opportunity criterion satisfies such a condition, as it can perfectly live with the fact that individuals have incoherent preferences. Indeed, it does not say anything on how individuals *should* choose. However, one may also like to formulate the reconciliation problem in a way that is about finding normative criteria that are *based on individual psychology*. This condition is of course stronger than the one formulated by McQuillin and Sugden (2012). But as for what is of the concern of the opportunity criterion, one may fairly ask whether the positive feeling of increasing choice is superior to the negative feeling of having more choice, all things considered. If the answer is positive, we may have good reason to use the opportunity criterion for normative analysis. But if the answer is negative, we may have good reason to give more importance to psychological processes, feeling, affects, etc. — anything related to how individuals perceive their situation — rather than to provide individuals with more opportunity to choose, *regardless what they would feel with more opportunity*. Note that the opportunity criterion states that even if individuals would be less happy with more opportunity, it would still justify enhancing individuals' opportunity to choose from. Yet many may find this principle misleading: is not what individuals *feel/perceive* that ultimately matters?

Third, how to measure opportunity (and whether it is measurable at all) is a complex debate in social choice that is far from being consensual. At first sight, this criticism may sound unfair given the individual maximisation-based criteria offered in the literature. Is measuring happiness consensual, given the many different interpretations 'happiness' can take? Is the measurement of true preference (or choice free from cognitive biases) consensual, knowing that it requires to assume that true preference exists beneath the psychology of individuals? Regardless the methodological issues associated with each normative criterion, it is however less obvious to debate on the measurement of pleasure and pain or on the measurement of true preference, at least when compared to the measurement of opportunity. In fact, how opportunity is to be measured is an enquiry on its own in social choice that is far from being consensual.[25] In a nutshell, there are at least three competing approaches offered in the literature.

- *Pure quantity*. Opportunity can simply be measured in terms of the number of alternatives contained in the opportunity (or choice) set. For example, to solve the problem that non-nested sets are not comparable, we may simply say that $O_1'' = \{r, s, t, u, v, x, y, z\}$ provides more opportunity than $O_2'' = \{w\}$ because O_1'' contains more alternatives than O_2''. Obviously, the issue with the pure quantity approach is that it is quite naive. It exclusively counts the number of alternatives without distinguishing the *nature* of these alternatives (Pattanaik and Xu 1990). To solve this problem, an alternative measure of opportunity could differentiate between the diversity of the alternatives.[26] For example, it may sound relatively consensual that the opportunity set $O_1''' = \{bluecar, redcar, greencar, yellowcar, blackcar, whitecar\}$ provides less opportunity than the opportunity set $O_2''' = \{blue\ car, bicyle, train\}$, simply because the combination of alternatives in O_2''' are more diversified than the

[25] See Pattanaik and Xu (1990), Sen (1991) and Sugden (2003, 1998, 2010) for a debate.

[26] The literature on elaborating a diversity-metric of opportunity includes Pattanaik and Xu (2000), Bossert, Pattanaik, and Xu (2003) and van Hees (2004), among others.

combination of alternatives in O_1'''.[27]

- *Potential preference*. Another measurement of opportunity is 'the range of preference that individuals might have had in relevant circumstances' (Sudgen 1998, p. 323). This approach is given support by Sen (1991), who argues that preference-satisfaction and freedom are very deeply interrelated. In this approach, opportunity-metric cannot be dissociated with what individuals would like to pursue, because it is specifically in being able to satisfy their preferences that individuals have more opportunity. According to Sugden (2010), this measurement of opportunity is however problematic because it inevitably associates potential preference with a conception of what individuals reasonably/morally would like to choose. In other words, potential preference requires to define what goodness objectively is — an enterprise that liberal tenants of the choice-based and opportunity criteria would like to stay away from.[28]

- *Opportunity without metric (mutual advantage)*. Yet another approach to opportunity endorsed by Sugden (2010) is that opportunity cannot be measured because it would require to objectively define what it is (a stance that the author is opposed to). In Sugden's (2010) words, 'opportunity is an open-ended concept: often, we cannot specify in concrete terms what a person does or does not have the opportunity to do, or what the value is of the things that she might do' (p. 48). Although opportunity is not measurable in this approach, the point of the author is that we can say, whether within a given economy, all feasible opportunities have been made available — and this is what ultimately counts in the author's conception of opportunity. The problem of leaving opportunity without measurement is however that it may disappoint some economists and philosophers, who would be reluctant to say that there is no objective characteristic associated with opportunity (such as pure quantity or diversity).

Fourth (and along what has been discussed previously), the metaphysical interpretation of responsibility is interpreted as an implicit axiom that one is required to accept. But one may easily argue that responsibility is not a characteristic that everyone is expected to have. The idea is that providing an individual with more opportunities is meaningless if such an individual is not responsible for her own choice, nor autonomous enough to make her own decisions. Consider for example students who are offered a course list, and because of their inexperience and youth cannot seriously be held responsible for choosing among the many available alternatives (Schwartz 2004 [2016], p. 18). Are individuals 'responsible/autonomous enough' to benefit from more choice? Sugden (2004, 2018) assumes so. But in situations where they cannot be expected to be, more opportunity

[27]In this example, some may still argue that defining opportunity through diversity yields to the philosophical problem that one cannot expect individuals to attribute the same value to the properties of available alternatives. For example, a rich individual who has a passion for car collections may give more opportunity value to O_1''' than to O_2'''. From a philosophical viewpoint, defending a diversity-metric of opportunity would maybe require to believe in public deliberation, i.e. that there are some alternative on which almost everyone would value the properties of available alternatives similarly, and that such collective judgement forms the objective truth. For a philosophical defence of public deliberation, see Sen (2009).

[28]Yet this problem may not be that problematic for pragmatic purposes. For example, one could reasonably consider that there are some values such as *human capabilities* (Nussbaum 2000) that convincingly define what it takes to have more opportunity for every human being (e.g. being able to vote, to go to school, to have access to medical care, etc.).

may be harmful. Consider limited cognitive abilities of individuals who face complex and opaque information. One cannot always expect individuals to be perfectly informed about what they choose. The capacity of being able to make *enlightened* choices is then a serious concern for the opportunity criterion, where the concept of responsibility only holds if individuals are already well informed, well experienced, etc. Education plays an essential role because it is not only a matter of having the 'right' type of information, but also a matter of *how* the information is conveyed. Yet this is an aspect neglected by Sugden (2004, 2018) in his defence of the opportunity criterion.

4 Solving the Reconciliation Problem: A Synthesis

The table below summarises the methodological and theoretical issues associated with each normative criterion previously discussed.

Table 1: Summary of the methodological and theoretical issues of the *experienced utility*, *true preference*, *choice-based* and *opportunity* criteria

			Methodological and theoretical issues
Individual Utility Maximisation-Based Criteria	**Happiness**	**Experienced utility**	Hedonism is narrow dimension of what makes the good life
			Takes a peculiar conception of happiness in which only moment utility matters
			Assumes interpersonal comparisons of utilities (particularly cardinality)
	Well-being	**True preference**	There is no psychological support for the existence of true preference
			Can only be justified under specific cases where distortions from cognitive biases make individuals worse-off
			The dominant policy recommendation based on this normative criterion (libertarian paternalism) struggles to combine paternalistic and liberal values
		Choice-basis	Based on the same concept than true preference, so the same critics apply
			Depends (paradoxically) on non-choice data
			Is (presumably) not ethically grounded at all
Freedom-Based criterion		**Opportunity**	Forbids to make comparisons between non-nested sets
			Choice overload and self-constraint may challenge the idea that more choice (or opportunity) is always better than less
			Measuring opportunity is a complex debate in social choice
			Requires to endorse the axiom that individuals are responsible and autonomous beings under all circumstances

Because of these issues of different nature, the framework I initially introduced based on the *general*, *ethical* and *practical* requirements is now helpful to assess the relevance of each normative criterion for normative analysis. As we shall see, if these requirements are accepted then no normative criterion satisfy all three requirements. In order to make the evaluation more informative, the *ethical requirement* can be separated in the following four questions and the *practical requirement* can be separated in the following two questions.

1. **General requirement.** *Can the normative criterion apply to a wide range of choice situations?* [theoretical and philosophical question]

2. **Ethical requirement**

    2.1. *Can the normative criterion capture the many different aspects of life that individuals can find valuable?* [ethical question]

    2.2. *Does the normative criterion actually capture the many different aspects of life that individuals can find valuable?* [ethical and empirical question]

    2.3. *Is the normative criterion supposed to be psychologically grounded?* [theoretical question]

    2.4. *Is the normative criterion actually psychologically grounded?* [empirical question]

3. **Practical requirement**

    3.1. *Does the normative criterion provide a measurement of individuals' states of affairs?* [theoretical question]

    3.2. *Is the measure relatively consensual?* [empirical question]

The table below provides a visual representation of the evaluation of the four normative criteria with respect to the *general*, *ethical* and *practical* requirements. It is the result of everything that has being said so far. The questions are formulated in a way that a positive answer ('YES') corresponds to an advantage, and a negative answer ('NO') corresponds to a disadvantage.

Table 2: Summary of whether the *experienced utility*, *true preference*, *choice-based* and *opportunity* criteria fulfil the *general, ethical* and *practical* requirements

		Experienced utility	**True preference**	**Choice-basis**	**Opportunity**
1. General requirement	*1.1. Can it apply to a wide range of choice situations?*	NO (only experiences of pain and pleasure)	NO (only when distortions from rationality make individuals worse off)	NO (only when distortions from rationality make individuals worse off)	NO (only nested sets)
2. Ethical requirement	*2.1. Can it capture the many different aspects of life that individuals can find valuable?*	NO (hedonism: narrow dimension of the good)	YES	YES	YES
	2.2. Does it actually capture the many different aspects of life that individuals can find valuable?	-	NO (struggles to preserve autonomy)	NO (presumably not ethically grounded)	NO (choice overload and self-constraint)
	2.3. Is it supposed to be psychologically grounded?	YES	YES	YES	NO
	2.4. Is it actually psychologically grounded?	YES	NO (inner rational agent critique)	NO (inner rational agent critique and depends on non-choice data)	-
3. Practical requirement	*3.1. Does it provide a measurement of individuals' states of affairs?*	YES	YES	YES	NO
	3.2. Is the measure relatively consensual?	YES	YES	YES	NO (complex debate in social choice)

As we can see, some normative criteria fulfil more requirements than others, but it would be misleading to count the approach that generates the more 'YES' as the one that should be preferred because some issues may be more concerning than other ones. For example, the experienced utility criterion fulfils relatively well *2.3., 2.4., 3.1.* and *3.2.*, but poorly *1.1.* and *2.1.*, which — in my view — are essential requirements for a 'good' normative criterion. On the other hand, recall that one may say that self-constraint only plays a small part in economic life (and perhaps in the scope of what generally matters to individuals), so we could almost respond a 'YES' to *2.2.* for the opportunity criterion if we consider choice overload not to be significantly supported by empirical evidence.

5 Discussion

With the growing interest of academic research in the reconciliation problem, reconciling normative and behavioural economics involves interdisciplinary thinking at the intersection of economics, psychology and philosophy. The central claim of this paper is that solving the reconciliation problem requires to have a normative criterion (i) that can apply to many choice situations, (ii) that can say something consistent about what makes individuals better off, and (iii) that allows for practically measuring their 'better off' states. In this survey, I propose a simple framework that serves as a basis to evaluate whether the normative criteria offered in behavioural economics reasonably satisfy these three requirements. The result is that none of them satisfies them all.

5.1 On the Difficulty of a 'Meta-Assessment'

At their current status, I acknowledge these requirements to be abstract representations that are not strictly delimitated, and that it is up to the reader to agree with the way I have defined these requirements. Indeed, the aim of this paper is first to make a critical survey of the reconciliation problem (which, as said previously, is to my knowledge inexistant), but the grounding or standard for 'good' would require further justification. In particular, one may rightfully wonder whose ideas about what is properly 'general,' 'ethical,' and 'practical' are being considered. Also, one may argue that an *ex-post* (instead of an *ex-ante*) proposition of these meta-criteria is preferable, as it would require less commitment to what we should consider as being 'general', 'ethical' and 'pratcial'. Instead of stating what we may think the characteristics of a 'good' normative criterion should be and then checking whether each normative criterion satisfies them all (the *ex-ante* approach), the *ex-post* approach consists, in a more transcendental way, to check what economists and philosophers consider to be the necessary features of a good normative criteria and then to extrapolate what they must consider to be the necessary features of a good normative criterion. In response to these two limitations (to which I am very much sympathetic), I advance the following two points that I hope would help stimulating the debate.

First, a critical review of each normative criterion like the one proposed in Section 3 (which obviously cannot be exhaustive) is useful in order to have an overview about what the most concerning methodological and theoretical problems of each normative criterion are. Shall we find out that the reconciliation problem cannot be solved because no normative criterion satisfies some important requirements that economists and philosophers judge to be essential (even if other than the three presently proposed), it would already be an interesting result (perhaps a disappointing one, but still an interesting result). Second, with the present survey I surely do not hope to resolve some methodological disagreements between leading experts in behavioural and normative economics, nor hope for a one-shot consensus among the community of researchers interested in the reconciliation problem. The point is, if some economists and philosophers do not even consensually agree on whether normative analysis should be either preference-based or choice-based, I surely cannot expect them to reach a consensus by anytime soon about how the reconciliation problem is best tackled. Yet this synthesising work has the merit of inviting economists and philosophers to debate about whether the three requirements I here propose are relatively consensual. If they are, we can at least advance on the other laborious task of finding a normative criterion that fulfils all three requirements.

Ultimately, the usefulness of the present work is that it can stimulate avenues of future research, mainly about proposing alternative normative criteria that are neither based on the *happiness*, nor the *well-being* nor the *freedom* interpretation of preference-satisfaction. It is important to remind ourselves that the three interpretations of *happiness*, *well-being* and *freedom* are only some among many ethical loci offered in ethics. Also, it is not because standard welfare economics is grounded on preference-satisfaction that we should necessarily continue in that direction.[29] The point is albeit happiness, well-being and freedom are the three main interpretations of what makes the good life in normative economics, the vast field of ethics has obviously many other theories to offer about what makes the good life. Economists interested in normative analysis may be interested in these alternative ethical theories, just as they may be interested in happiness, well-being and freedom. I here conclude by briefly exposing two other ethical interpretations that are unfamiliar with the tradition of normative economics but which are, at their current status, still at an embryonic stage: *virtue ethics* and *meaning*.

5.2 Other Ethical-Loci-Based Criteria

5.2.1 Virtue Ethics

An alternative normative approach that is worth being considered as a solution to the reconciliation problem is the *virtue ethics* approach proposed by Bath, Ogaki and Yaguchi (2015, 2017) and synthesised in Ogaki and Tanaka (2017, Ch. 11).[30] The authors start from the following tridimensional taxonomy of the 'relevant' ethical theories to normative economics: consequentialism (e.g. welfarism or utilitarianism), deontology (e.g. egalitarianism) and virtue ethics, which they note to be largely ignored in normative economics. The authors argue that public policy is mainly a matter of meta-preferences over social states of affairs, such as preferences for e.g. less addiction and more patience. In their words, 'it is virtuous to be pleased about what we should be pleased about' (Ogaki and Tanaka 2017, p. 193). The authors consider meta-preferences to be

[29] Anecdotally, the strongest *status quo* bias ever observed in behavioural economics may not be in subject's behaviour, but (ironically) in economists' behaviour. See Davis (2016), who provides a sociology-of-science analysis about the practice of economists towards their own science. The author argues that economists often tend to experience strong loss aversion and *status quo* bias in the assumptions they make.

[30] See Bruni and Sugden (2013) for a philosophical defence that the market economy can be seen as a sphere of virtue.

explicit statements of what makes the good life, e.g. the ones implicitly endorsed by Thaler and Sunstein (2009) such as 'smoking is bad' or 'eating healthy is good'. The novelty of Bath, Ogaki and Yaguchi (2015, 2017) is to extend welfarism by adding virtue ethics as a component of the social evaluation. That is to say, they propose that social evaluation should not solely depend on a standard welfare function $W(u_1(x), ..., u_n(x))$, but also on what they call a 'moral evaluation function' $M(\psi_1(x), ..., \psi_n(x); \psi^*)$, where x is a social state, $u_i(x)$ the utility function of individual i, $\psi_i(x)$ a function that expresses properties of the endogenous utility function of individual i and ψ^* the ethical benchmark of the idealistic social good. For example, ψ^* can represent zero addiction, full patience or any other relevant ethical benchmark the society judges to be good based on a certain ideal of what it means to live a 'virtuous' life. We can then express the ethical premise of the *virtue ethics* criterion as follows.

Virtue ethics (ethical premise). *It is good to satisfy individuals' meta-preferences over what is judged to be desirable for society.*

The virtue ethics criterion can be interpreted as a variant of the true preference criterion, where (i) there exists an ethical benchmark of the idealistic social good ψ^* and (ii) the set of psychological states is not entirely a matter of *subjective* evaluation (what does each individual judge to be her own good?) but also a matter of *objective* evaluation (what does society judge to be individuals' good?). The 'social objective function' O is then a function of W and M. That is, $O(W(x), M(x))$ is a function that evaluates social states by considering both welfarism and virtue to be what constitutes goodness. The idea is that individuals can surely judge what their own good is, but there are also some things in life such as patience or being healthy that are 'ethically desirable' for everyone. *Who* judges what is ethically desirable for individuals? The social planner (again), as this is necessarily implied by the social objective function O. However, it is not excluded that the social planner is here only a representative of the individual members of the society who consensually agree on what the social good is.

From an ethical viewpoint, the main limit of this approach is that it tries to combine two theories of ethics that seem hardly compatible: consequentialism (here welfarism) and virtue ethics. The former is particular to the tradition of welfare economics and social choice. According to welfarism, the social good is determined by (and only by) the well-being of individuals (Sen 1977). However, virtue ethics is supposed to entail an informational basis that goes 'beyond' well-being, such as patience and no addiction (in the authors' own terms). Thus, we have, on the one hand, goodness defined in terms of well-being, and on the other hand, goodness defined in terms of acquired character-traits or dispositions that are judged to be good (e.g. patience or no addiction). The question is, how are we to combine both of these ethical theories into a normative criterion? The approach of the authors seems a bit challenging, as one may notice that (i) they define $\psi_i(x)$ as a function that expresses properties of the endogenous *utility function* of individual i (that is, virtue is nothing more than a component that is already included in one's utility) and (ii) they do not say anything about when virtue can outweigh purely welfarist considerations (that is, trade-offs between preference-satisfaction and virtue remain unspecified). But is the goal not to introduce virtue ethics in order to *replace* welfarism, or at least, to take an alternative ethical locus to individual and social evaluation *other than* happiness, well-being or freedom? Unfortunately, the virtue ethics criterion proposed by Bath, Ogaki and Yaguchi (2015, 2017) is ultimately a component of the generalised social welfare function O. That function can merely be seen as an extension of welfarism but not as an alternative that departs from welfarism.

5.2.2 Meaning

There is however another possible ethical locus for normative economics that is even less known in the literature: the *meaning* individuals give to the realisation of their states of affairs. This approach is given support in Loewenstein (1999), Karlsson, Loewenstein, and McCafferty (2004) and Dold and Stanton (2021). The aim of Loewenstein (1999) is to 'enrich' the notion of utility by incorporating the dimension of non-consumption into the utility function. The author particularly does so by taking the case of human activity that does not derive from pleasure. As Loewenstein (1999) argues,

> 'Despite the blossoming of the utility concept and expanding appreciation for the diverse determinants of utility, the list of human motives that have been codified in utility functions, and hence incorporated into economic analyses, remains seriously incomplete.' (p. 316)

Loewenstein (1999) takes mountaineering as an illustrative case, where individuals who undertake such activity do so for reasons that are not obvious from an external standpoint. Indeed, the experience of mountaineering is very often miserable. Yet the aim is specifically to account for these unilluminating reasons because they appear to have the most sense. If individuals undertake miserable activities such as mountaineering, learning the theory of music, or writing a PhD thesis, it must be something other than happiness (or well-being, or freedom) that motivates them to choose or to experience something. Since the *meaning* criterion could perhaps entail the broadest range of what matters to individuals, it could potentially better satisfy the general and ethical requirements than any other normative criterion here reviewed.

One reason that meaning would be of good use for normative analysis can be found in Davis (2011), who fairly argues that economic choices intimately relate to our human conditions. The argument is that the economic part of one's decision (whatever an 'economic choice' might mean) has become tremendously important in individuals' lives, to the point that it is on choices regarding mortgage, where to live, saving plans, etc., that individuals plan their lives as a long-term process. Normative economics typically takes preference-satisfaction as its normative standard, but is silent about what hides behind consumption and individual behaviour. Instead, a broader perspective normative economics could take is to consider every choice as being meaningful — but with different degrees. For example, individuals can buy a bungie dive ticket as a leisure activity, they can buy food at the supermarket, invest on financial markets, etc., and yet attribute a *meaning* behind all these decisions, *even if they do not derive any pleasure or satisfaction from consuming these goods or undertaking these activities*. The ethical premise of the meaning criterion would then be expressed as follows.

Meaning (ethical premise). *It is good to realise individuals' expectations about living a meaningful life.*

The particularity of meaning is that it belongs to a category other than happiness, well-being, freedom and virtue ethics. It is an ethical locus of its own kind that can be traced back in *existentialism*: a philosophical tradition that is unfamiliar with normative economics. Further research could then aim to (i) argue more extensively why such normative criterion would matter, as in Karlsson, Loewenstein, and McCafferty (2004) and Dold and Stanton (2021); (ii) formalise a model of preference formation in order to understand the psychological mechanism under which individuals choose according to what they believe is meaningful for them; (iii) argue that economic policy has good interest in focusing on enhancing the desirability of individuals to engage in meaningful choices as a source of life fulfilment. As a consequence, the aim of public policy would be to enhance meaningful choices or experiences in life, *even if these choices or experiences provide less happiness, well-being and freedom/autonomy*.

6 Concluding Remarks

With all the normative criteria that have been surveyed in the present article (*experienced utility*, *true preference*, *choice-basis*, *opportunity*, *virtue ethics* and *meaning*), I hope to have provided economists and philosophers with a rich and concise literature review that will stimulate promising perspectives of research about the ambitious challenge of 'reconciling' normative and behavioural economics. I now hand it over to the reader.

References

Arneson, R. J. (1990). Liberalism, distributive subjectivism, and equal opportunity for welfare. *Philosophy and Public Affairs 19*(2), 158–194.

Benartzi, S. and R. H. Thaler (2002). How much is investor autonomy worth? *The Journal of Finance 57*(4), 1593–1616.

Bentham, J. (2007). *An Introduction to the Principles of Morals and Legislation*. Dover Philosophical Classics.

Berg, N. (2003). Normative behavioral economics. *The Journal of Socio-Economics 32*(4), 411–427.

Berg, N. and G. Gigerenzer (2010). As-if behavioral economics: neoclassical economics in disguise? SSRN Scholarly Paper, Social Science Research Network.

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association 7*(2-3), 267–319.

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

Bernheim, B. D. and A. Rangel (2004). Addiction and cue-triggered decision processes. *American Economic Review 94*(5), 1558–1590.

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

Bernheim, B. D. and A. Rangel (2008). Choice-theoretic foundations for behavioral welfare economics. In A. Caplin and A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*, pp. 155–192. Oxford University Press.

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

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2008). How are preferences revealed? *Journal of Public Economics 92*(8-9), 1787–1794.

Bhatt, V., M. Ogaki, and Y. Yaguchi (2015). Normative behavioural economics based on unconditional love and moral virtue. *The Japanese Economic Review 66*(2), 226–246.

Bhatt, V., M. Ogaki, and Y. Yaguchi (2017). Introducing virtue ethics into normative economics for models with endogenous preferences. Rochester Center for Economic Research Working Paper N° 600.

Bleichrodt, H., J. L. Pinto, and P. P. Wakker (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science 47*(11), 1498–1514.

Bossert, W., P. K. Pattanaik, and Y. Xu (2003). Similarity of options and the measurement of diversity. *Journal of Theoretical Politics 15*, 405–421.

Botti, S. and S. S. Iyengar (2006). The dark side of choice: when choice impairs social welfare. *Journal of Public Policy and Marketing 25*(1), 24–38.

Bruni, L. and R. Sugden (2013). Reclaiming virtue ethics for economics. *Journal of Economic Perspectives 27*(4), 141–164.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue, and M. Rabin (2003). Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism". *University of Pennsylvania Law Review 151*(3), 1211–1254.

Chernev, A., U. Böckenholt, and J. Goodman (2015). Choice overload: a conceptual review and meta-analysis. *Journal of Consumer Psychology 25*(2), 333–358.

Dalton, P. S. and S. Ghosal (2011). Behavioral decisions and policy. *CESifo Economic Studies 57*(4), 560–580.

Dalton, P. S. and S. Ghosal (2012). Decisions with endogenous frames. *Social Choice and Welfare 38*, 585–600.

Davis, J. B. (2011). *Individuals and Identity in Economics*. Cambridge University Press.

Davis, J. B. (2016). Economists' odd stand on the positive-normative distinction: a behavioral economics view. In G. F. DeMartino and D. McCloskey (Eds.), *The Oxford Handbook of Professional Economic Ethics*, pp. 199–218. Oxford University Press.

DellaVigna, S. (2009). Psychology and economics: evidence from the field. *Journal of Economic Literature 47*(2), 315–372.

Dhami, S. S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press.

Dolan, P. and D. Kahneman (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal 118*(525), 215–234.

Dold, M. F. and C. Schubert (2018). Toward a behavioral foundation of normative economics. *Review of Behavioral Economics 5*(3-4), 221–241.

Dold, M. F. and A. Stanton (2021). I choose for myself, therefore I am: the contours of existentialist welfare economics (forthcoming). *Erasmus Journal for Philosophy and Economics*.

Edgeworth, F. Y. (1881). *Mathematical Psychics*. C. Kegan Paul & Co.

Elster, J. (1998). *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge University Press.

Elster, J. (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press.

Elster, J. (2016). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.

Fine, B. (1995). On the relationship between true preference and actual choice. *Social Choice and Welfare 12,* 353–361.

Fredrickson, B. L. and D. Kahneman (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology 65*, 45–55.

Gauthier, D. (1986). *Morals by Agreement*. Clarendon Press.

Goodin, R. (1992). Laundering Preferences. In J. Elster and A. Hylland (Eds.), *Foundations of Social Choice Theory* (reprinted ed.)., pp. 75–101. Cambridge University Press.

Grüne-Yanoff, T. (2012). Old wine in new casks: libertarian paternalism still violates liberal principles. *Social Choice and Welfare 38*(4), 635–645.

Guala, F. and L. Mittone (2015). A political justification of nudging. *Review of Philosophy and Psychology 6*(3), 385–395.

Halpern, D. (2015). *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. Allen.

Hands, D. W. (2014). Normative ecological rationality: normative rationality in the fast-and-frugal-heuristics research program. *Journal of Economic Methodology 21*(4), 396–410.

Hands, D. W. (2020). Libertarian paternalism: taking Econs seriously. *International Review of Economics*, 1–23.

Harsanyi, J. C. (1977). Rule utilitarianism and decision theory. *Erkenntnis 11*(1), 25–53.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press.

Hausman, D. M. (2016). On the Econ within. *Journal of Economic Methodology 23*(1), 26–32.

Hausman, D. M. (2018). The bond between positive and normative economics. *Revue d'Économie Politique 128*(2), 191–208.

Hausman, D. M., M. McPherson, and D. Satz (2016). *Economic Analysis, Moral Philosophy, and Public Policy* (third ed.). Cambridge University Press.

Hédoin, C. (2015). From utilitarianism to paternalism: when behavioral economics meets moral philosophy. *Revue de Philosophie Économique 16*(2), 73–106.

Hédoin, C. (2017). Normative economics and paternalism: the problem with the preference-satisfaction account of welfare. *Constitutional Political Economy 28*(3), 286–310.

Heidhues, P., J. Johnen, and B. Köszegi (2020). Browsing versus studying: a promarket case for regulation (forthcoming). *Review of Economic Studies*.

Hutchinson, J. M. C. (2005). Is more choice always desirable? Evidence and arguments from leks, food selection, and environmental enrichment. *Biological Reviews 80*, 73–92.

Infante, G., G. Lecouteux, and R. Sugden (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology 23*(1), 1–25.

Iyengar, S. S. (2010). *The Art of Choosing*. Hachette Digital.

Iyengar, S. S. and M. R. Lepper (2000). When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology 79*(6), 995–1006.

Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft 150*(1), 18–36.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, and N. Schwarz (Eds.), *Well-being: The Foundations of Hedonic Psychology*, pp. 3–25. Russell Sage Foundation.

Kahneman, D. (2000). Experienced utility and objective happiness: a moment-based approach. In D. Kahneman and A. Tversky (Eds.), *Choices, Values, and Frames*, pp. 673–692. Cambridge University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books.

Kahneman, D. (2018). Why Nobel prize winner Daniel Kahneman gave up on happiness (interview). `https://www.haaretz.com/israel-news/.premium.MAGAZINE-why-nobel-prize-winner-daniel-kahneman-gave-up-on-happiness-1.6528513`. [Online; accessed 01-August-2020].

Kahneman, D., B. L. Fredrickson, C. A. Schreiber, and D. A. Redelmeier (1993). When more pain is preferred to less: adding a better end. *Psychological Science 4*(6), 401–405.

Kahneman, D. and A. B. Krueger (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives 20*(1), 3–24.

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science 306*(5702), 1776–1780.

Kahneman, D. and J. Snell (1990). Predicting utility. In R. M. Hogarth (Ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, pp. 295–310. University of Chicago Press.

Kahneman, D. and J. S. Snell (1992). Predicting a changing taste: do people know what they will like? *Journal of Behavioral Decision Making 5*(3), 187–200.

Kahneman, D. and R. Sugden (2005). Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics 32*(1), 161–181.

Kahneman, D. and R. H. Thaler (2006). Anomalies: utility maximization and experienced utility. *Journal of Economic Perspectives 20*(1), 221–234.

Kahneman, D. and A. Tversky (Eds.) (2000). *Choices, Values, and Frames*. Cambridge University Press.

Kahneman, D. and C. Varey (1991). Notes on the psychology of utility. In J. Elster and J. E. Roemer (Eds.), *Interpersonal Comparisons of Well-Being*, pp. 127–163. Cambridge University Press.

Kahneman, D., P. P. Wakker, and R. Sarin (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics 112*(2), 375–406.

Karlsson, N., G. Loewenstein, and J. McCafferty (2004). The economics of meaning. *Nordic Journal of Political Economy 30*, 61–75.

Köszegi, B. and M. Rabin (2007). Mistakes in choice-based welfare analysis. *American Economic Review 97*(2), 477–481.

Köszegi, B. and M. Rabin (2008). Revealed mistakes and revealed preferences. In A. Caplin and A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*, pp. 193–209. Oxford University Press.

Lecouteux, G. (2015). In search of lost nudges. *Review of Philosophy and Psychology 6*(3), 397–408.

Lecouteux, G. (2016). From homo economicus to homo psychologicus: the Paretian foundations of behavioural paternalism. *Œconomia. History, Methodology, Philosophy 6*(2), 175–200.

Loewenstein, G. (1999). Because it is there: the challenge of mountaineering ... for utility theory. *Kyklos 52*(3), 315–343.

Loewenstein, G. and E. Haisley (2008). The economist as therapist: methodological ramifications of "light" paternalism. In A. Caplin and A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*, pp. 210–245. Oxford University Press.

Loewenstein, G. and P. A. Ubel (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics 92*(8-9), 1795–1810.

Madrian, B. C. and D. F. Shea (2001). The power of suggestion: inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics 116*(4), 1149–1187.

Manzini, P. and M. Mariotti (2014). Welfare economics and bounded rationality: the case for model-based approaches. *Journal of Economic Methodology 21*(4), 343–360.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012). Revealed attention. *American Economic Review 102*(5), 2183–2205.

McQuillin, B. and R. Sugden (2012). Reconciling normative and behavioural economics: the problems to be solved. *Social Choice and Welfare 38*(4), 553–567.

Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review 99*(3).

Mitrouchev, I. (2020). Back to Aristotle? Explorations of objective happiness. Working Paper hal-02915805.

Mitrouchev, I. and V. Buonomo (2020). Identity, personal persistence and normative economics. Working Paper hal-02903633.

Nussbaum, M. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge University Press.

Ogaki, M. and S. C. Tanaka (2017). *Behavioral Economics: Toward a New Economics by Integration with Traditional Economics*. Springer.

Pattanaik, P. K. and Y. Xu (1990). On ranking opportunity sets in terms of freedom of choice. *Recherches Économiques de Louvain (Louvain Economic Review) 56*(3-4), 383–390.

Pattanaik, P. K. and Y. Xu (2000). On diversity and freedom of choice. *Mathematical Social Sciences 40*, 123–130.

Pinto-Prades, J.-L. and J.-M. Abellan-Perpiñan (2012). When normative and descriptive diverge: how to bridge the difference. *Social Choice and Welfare 38*(4), 569–584.

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In K. J. Arrow, E. Colombatto, M. Perlaman, and C. Schmidt (Eds.), *The Rational Foundations of Economic Behaviour*, pp. 225–250. McMillian.

Railton, P. (2003). *Facts, Values, and Norms: Essays toward a Morality of Consequence*. Cambridge University Press.

Rebonato, R. (2012). *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Palgrave Macmillan UK.

Rebonato, R. (2014). A critical assessment of libertarian paternalism. *Journal of Consumer Policy 37*(3), 357–396.

Redelmeier, D. A. and D. Kahneman (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain 66*(1), 3–8.

Redelmeier, D. A., J. Katz, and D. Kahneman (2003). Memories of colonoscopy: a randomized trial. *Pain 104*(1), 187–194.

Reisch, L. A. and C. R. Sunstein (2016). Do europeans like nudges? *Judgment and Decision Making 11*(4), 310–325.

Reisch, L. A., C. R. Sunstein, and W. Gwozdz (2017). Beyond carrots and sticks: Europeans support health nudges. *Food Policy 69*, 1–10.

Rizzo, M. J. and D. G. Whitman (2009). The knowledge problem of new paternalism. *BYU Law Review* (4), 905–968.

Rizzo, M. J. and D. G. Whitman (2019). *Escaping Paternalism*. Cambridge University Press.

Rubinstein, A. and Y. Salant (2012). Eliciting welfare preferences from behavioural data sets. *The Review of Economic Studies 79*(1), 375–387.

Salant, Y. and A. Rubinstein (2008). (A, f): choice with frames. *The Review of Economic Studies 75*(4), 1287–1296.

Scheibehenne, B. (2008). *The Effect of Having Too Much Choice*. PhD thesis. Humboldt-Universität zu Berlin.

Scheibehenne, B., R. Greifeneder, and P. M. Todd (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research 37*(3), 409–425.

Schreiber, C. A. and D. Kahneman (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General 129*(1), 27–42.

Schwartz, B. (2016). *The Paradox of Choice* (revised ed.). HarperCollins.

Scoccia, D. (2019). Paternalisms and nudges. *Economics and Philosophy 35*(1), 79–102.

Sen, A. (1977). On weights and measures: informational constraints in social welfare analysis. *Econometrica 45*(7), 1539–1572.

Sen, A. (1991). Welfare, preference and freedom. *Journal of Econometrics 50*(1-2), 15–29.

Sen, A. (2009). *The Idea of Justice*. Harvard University Press.

Sudgen, R. (1998). The metric of opportunity. *Economics and Philosophy 14*(2), 307–337.

Sugden, R. (2003). Opportunity as a space for individuality: its value and the impossibility of measuring it. *Ethics 113*(4), 783–809.

Sugden, R. (2004). The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American economic review 94*(4), 1014–1033.

Sugden, R. (2007). The value of opportunities over time when preferences are unstable. *Social Choice and Welfare 29*(4), 665–682.

Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy 19*(3), 226–248.

Sugden, R. (2010). Opportunity as mutual advantage. *Economics and Philosophy 26*(1), 47–68.

Sugden, R. (2015). Looking for a psychology for the inner rational agent. *Social Theory and Practice 41*(4), 579–598.

Sugden, R. (2017a). Characterising competitive equilibrium in terms of opportunity. *Social Choice and Welfare 48*(3), 487–503.

Sugden, R. (2017b). Do people really want to be nudged towards healthy lifestyles? *International Review of Economics 64*(2), 113–123.

Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.

Sunstein, C. R. (2019). *On Freedom*. Princeton University Press.

Sunstein, C. R., L. A. Reisch, and M. Kaiser (2019). Trusting nudges? Lessons from an international survey. *Journal of European Public Policy 26*(10), 1417–1443.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization 1*(1), 39–60.

Thaler, R. H. (2018). From cashews to nudges: the evolution of behavioral economics. *American Economic Review 108*(6), 1265–1287.

Thaler, R. H. and S. Benartzi (2004). Save more tomorrow™: using behavioral economics to increase employee saving. *Journal of Political Economy 112*(S1), S164–S187.

Thaler, R. H. and C. R. Sunstein (2003). Libertarian paternalism. *American Economic Review 93*(2), 175–179.

Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness* (revised and expanded ed.). Penguin Books.

van Hees, M. (2004). Freedom of choice and diversity of options: some difficulties. *Social Choice and Welfare 22*, 253–266.

Varey, C. A. and D. Kahneman (1992). Experiences extended across time: evaluation of moments and episodes. *Journal of Behavioral Decision Making 5*(3), 169–185.

Welch, B. and D. Hausman (2010). To nudge or not to nudge? *Journal of Political Philosophy 18*(1), 123–136.

Whitman, D. G. and M. J. Rizzo (2015). The problematic welfare standards of behavioral paternalism. *Review of Philosophy and Psychology 6*(3), 409–425.