

Normative and Behavioural Economics: A Historical and Methodological Review

Ivan Mitrouchev

IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France

Accepted version: 01/02/2024

The European Journal of The History of Economic Thought

Abstract: Behavioural economics has not only posed serious challenges for the empirical adequacy of rational choice, but also for its normative status. Since the 1990s, a large body of work has proposed various normative approaches that account for inconsistent choices. Focusing on the relevant approaches, history, methods, and limits, this article offers a literature review of the relationship between normative and behavioural economics.

Keywords: choice – opportunity – preference – rationality – welfare

JEL codes: B21, B41, D60, D90

Acknowledgement: This research is supported by a grant from the Métropole Européenne de Lille (MEL). 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 am grateful to Francesco Guala, Cyril Hédoïn and Robert Sugden for helpful remarks, and most particularly to Wade Hands for insightful discussions that considerably helped me improving the ideas presented in this paper. Finally, I thank Pierre Van Zyl for proofreading and two anonymous referees for their extremely valuable remarks. All mistakes remain my own.

1. Introduction

Over the past few decades, behavioural economics – more particularly the heuristics-and-biases programme – has documented many psychological biases, which support the view that choices cannot easily be rationalised along the lines of traditional economic rationality.¹ These findings are not only concerning for descriptive (or positive) economics, which for around three quarters of a century has been based on the assumption of rational choice. They also have an impact on normative economics, which typically considers principles of rational choice as what indicates or constitutes individual welfare.² In particular, if (i) there is recurrent evidence that individuals do not behave according to the principles of rational choice, and if (ii) rational choice is supposed to indicate or constitute individual welfare, then (iii) it is not clear anymore what aspect of individual behaviour should identify what makes individuals better off.

Since the beginning of the 1990s, a consequent number of influential behavioural economists (among them Daniel Kahneman and Richard Thaler) has focused on alternative approaches to make normative analysis when individuals deviate from the principles of rational choice. Among the most prominent approaches, perhaps the first attempt is the normative programme of ‘Kahneman et al.’ (Kahneman 1994; Kahneman, Wakker, and Sarin 1997; among others), which proposed to evaluate welfare by measuring individuals’ level of hedonic experience. Then, ‘asymmetric’ and ‘libertarian’ paternalisms advocated by Thaler and Sunstein (2003) and Camerer et al. (2003) have been proposed to increase social welfare by exploiting individuals’ biases. Shortly afterwards, other approaches have proposed theoretical frameworks to infer welfare from observed choices when those are not necessarily consistent with the principles of rational choice (Bernheim and Rangel 2007, 2009; Salant and Rubinstein 2008; among others). Yet other approaches have aimed to combine theoretical, experimental and philosophical perspectives in the ambitious challenge of reconstructing normative economics with an opportunity-based criterion (Sugden 2018). Alternatively, a critical literature has been devoted to identify and discuss some issues of these approaches (Lecouteux 2015; Infante, Lecouteux, and Sugden 2016; Dold and Rizzo 2021; Thoma 2021; Grüne-Yanoff 2022; among many others), which has led some of the contributing authors to respond to those issues (Sunstein 2015; Bernheim 2021; Sugden 2021).

This article is (to my knowledge) the first attempt to extensively review the literature on the various propositions to make normative evaluation when individuals do not conform to rational choice. The contribution is twofold. I aim to put this body of literature into a (i) historical and (ii) methodological perspective. Regarding the historical perspective, most of this literature appeared in the last few decades, with perhaps the first attempt in behavioural economics to disentangle choice from welfare by March (1978), then taken back by Kahneman, Wakker, and Sarin (1997). However, methodological debates about what role psychology has played in economics, and the

¹ Rabin (1998) and DellaVigna (2009) offer extensive reviews. For historical reviews of behavioural economics, see Heukelom (2014) and Kao and Velupillai (2015).

² See Sugden (1991) for a survey of rational choice from a normative perspective – i.e. how individuals *ought to* choose.

reasons that economics departed from hedonic psychology are old.³ Framing this literature in a historical perspective (Section 2) is valuable in at least two ways. First, it is helpful to enlighten where and how this relatively recent literature originated from. Second, it is helpful to understand the historical influences of each contemporary approach, so that each can be meaningfully compared and discussed. With this respect, the historical analysis I provide in the first part sets the stage for the methodological review, which I address in the second part. Despite the many propositions on how to make normative evaluation consistent with behavioural economics, we still lack a general framework that aims at reviewing and systemising the related approaches. I review the main normative criteria proposed (Section 3), and suggest a multi-requirement framework to compare these normative criteria (Section 4). This is helpful to categorise the most concerning limits of these normative criteria, so that we can advance on how to overcome these limits.

The usual disclaimer about a literature review applies here. I do not mean to be exhaustive in covering *all* relevant aspects of the literature, as well as *every* limit associated with each normative approach. Instead, the aim is to offer a comprehensive survey, which is broad enough to cover some salient problems in this literature.

2. Historical Review

2.1. From Revealed Preference to Welfare

Standard economic theory relies on individual observable choices to analyse various phenomena, typically market mechanisms. With the concept of *revealed preference* formulated by Samuelson (1948), preferences can be revealed by observed choices as long as individuals satisfy the Weak Axiom of Revealed Preference (WARP).⁴ The theory of revealed preference was initially proposed as a method to derive demand functions in consumer choice. I here avoid making a distinction between different approaches of revealed preference theory that evolved in the twentieth century and rather refer to the generalisation formulated by Arrow (1959), where choices are broadened to include all finite sets of alternatives – meaning that the primitives are not individual demand functions but *choices*.⁵ That is, the individual (consumer) chooses between different alternatives (bundles of goods) which are available (affordable). According to this general formulation of revealed preference theory, if an individual chooses alternative x when y is available, he ‘reveals’ that he prefers x to y and will never choose y when x is available.

The development of revealed preference theory, coupled with ordinal utility theory promoted at the beginning of the twentieth century by Pareto (1909), Slutsky (1915), Hicks and Allen (1934), and then lately joined together by Houthakker (1950), characterised the general epistemological position taken by economists to make

³ On the historical relationship between psychology and economics, see Hands (2010). On the historical debate around the measurement of utility – from early neoclassical economics to behavioural economics – see Moscati (2018).

⁴ The axiom was later reformulated by Houthakker (1950) into the Strong Axiom of Revealed Preference (SARP) and by Afriat (1967) into the General Axiom of Revealed Preference (GARP).

⁵ For a detailed panorama of how different branches of revealed preference theory evolved over time, see Hands (2013).

choice theory a discipline that is free of any psychological content.⁶ A large part of the early historical debate was about the *utility* concept: whether it should be given a psychological interpretation in terms of *cardinality* and *hedonism* – following the utilitarian school of thought (Jevons 1871; Edgeworth 1881) – or whether it should be taken as a preference index, which in turn is nothing more than a ranking of alternatives – thus following the ordinalist school of thought (Robbins 1932; Hicks and Allen 1934), as well as the Samuelsonian (1938) school of thought, whose aim was to abandon the concept of *utility* at all. The arguments promoted by the ordinalist school of thought against the utilitarian school of thought can be resumed as follows. Giving utility a psychological interpretation is scientifically meaningless because individual psychology is empirically not observable (behaviourist/positivist argument), and *cardinality* and *interpersonal comparison of utilities* are not necessary features for descriptive demand theory (Occam's razor argument).

In the contemporary representation of utility – which is largely influenced by the ordinalist school of thought and revealed preference theory – individual choice provides *all* the necessary information to infer individuals' preferences (and therefore utility) of outcomes. This, however, holds on a strong presupposition of human behaviour: that individuals satisfy the principles of rational choice. In this sense, if GARP is taken as the benchmark of rationality (i.e. the 'reference point' from which deviations are to be measured) and is then subject to empirical test, it would not be surprising that choice theory would be concerned about a very restricted set of individual behaviour.⁷ Not only the standard model assumes that individuals (i) satisfy the principles of rational choices, and thus can be said to exhibit 'rational' preferences, but that (ii) individuals *maximise* these preferences.⁸ Tenants of the standard model of rational choice may fairly argue that the theory does not assume that individuals *actually* behave this way, but that it is concerned with individuals *who* satisfy these principles of rational choice. They may also argue that even if they do not satisfy these principles, the theory can be interpreted as being *normative* (what individuals *ought* to do) instead of being *descriptive* (what individuals *actually* do).⁹

My focus here is on the *normative* aspect of revealed preference theory. The axiomatic approach of revealed preference theory is the standard approach to derive utility from observed choices. But it is also the standard approach to evaluate *welfare*, where welfare is merely equal to choice. Precisely, since observed choices are supposed to

⁶ See in particular the historical interpretation of Hands (2010) about the role of psychology in consumer choice theory, from the early utilitarian stage of neoclassical economics to the ordinalist revolution.

⁷ This point was acknowledged by Paul Samuelson himself in a correspondence with Hendrik Houthakker regarding SARP: 'I imagine that you are right that many individuals looking at this paper will be induced to believe that there is after all very little, and very little of interest, in the modern theory of consumer's behavior. However, if this is indeed the truth, we should not try to keep it a secret. By all means let us make clear how little and how much the existing theories of economics contain.' (Paul Samuelson to Hendrik Houthakker on the 31st of July 1952). Few decades later, goodness-of-fit methods were proposed to measure how far individuals deviate from GARP (Afriat 1973; Houtman and Maks 1985; Varian 1990).

⁸ We can also include a third assumption, according to which (iii) individuals' subjective beliefs are updated using Bayes' rule. Yet in the contemporary approach of revealed preference I here discuss, such an assumption is (to my knowledge) absent of the theory.

⁹ A famous example in choice under uncertainty is Savage's (1954) reconsideration of the status of expected utility theory after he violated the *independence* axiom himself (Allais 1953). Savage's response to Allais was that the theory should not be seen as *descriptive* (how individuals actually choose) but as *normative* (how individuals should choose).

satisfy the principles of rationality in the standard framework, and since it is assumed that individuals maximise these preferences, there is no need to invoke any other criterion to indicate what makes individuals better off than observed choices. In this matter, *revealed preference*, *rational choice* and *welfare* are all conflated in the standard framework (Hausman 2012).¹⁰ Separating descriptive from normative analysis is thus straightforward in this framework. While the two are based on the same assumptions of individual behaviour, the latter is only concerned by analysing and comparing different situations without providing any answer to ethical questions such as what makes individuals better off. More generally, it is the approach to be found in some of the most influential microeconomic textbooks (Varian 1987 [2014]; Mas-Colell, Whinston, and Green 1995). These typically begin with consumer theory and then address questions of comparisons between different situations with the Pareto criterion, as well as demonstrating the virtues of the competitive market with the two theorems of welfare economics. Yet assessing different situations with the Pareto criterion and holding the two theorems of welfare economics not only presupposes that the hypotheses of perfect competition are met, but that the assumption of individual rational choice (which is not a weaker assumption) is met.

Since revealed preference theory treats with perfectly rational individuals who satisfy GARP, the possibility that they make a *mistake* in their choice is merely outside the scope of the theory. A couple of years after the development of revealed preference theory in the 1950s-1960s, the heuristics-and-biases programme has shaken a lot of rationality assumptions commonly endorsed in the standard framework. In a series of seminal contributions (Tversky and Kahneman 1973, 1974, 1981; Kahneman and Tversky 1979; among others), the heuristics-and-biases programme, developed in the 1970s, sought to (i) explore how heuristics lead to errors of judgement over objective probability, (ii) collect consistent and recurrent empirical findings that individuals deviate from the standard axioms of rational choice, and (iii) propose a new axiomatic approach to describe/explain/predict choice from the deviations of standard decision theory. At first, the interpretation of rationality deviations in the heuristics-and-biases programme referred to possible mistakes individuals would like to correct, had they been given an *ex-post* feedback of their choice. As Kahneman and Tversky (1979) put it:

'These departures from expected utility theory must lead to normatively unacceptable consequences, such as inconsistencies, intransitivities, and violations of dominance. Such anomalies of preference are normally corrected by the decision maker when he realizes that his preferences are inconsistent, intransitive, or inadmissible. In many situations, however, the decision maker does not have the opportunity to discover that his preferences could violate decision rules that he wishes to obey.' (Kahneman and Tversky 1979, p. 277)

This interpretation implies at least two assumptions commonly shared between revealed preference theory and the heuristics-and-biases programme. First, individuals have *ex-ante* and well-defined preferences (i.e. prior to their choices). Second, rationality principles provide guidance on what one *ought to do*. As the heuristics-and-biases programme provided evidence of systematic deviations from rational choice by categorising various biases in observed behaviour, there was a need to give those biases a normative interpretation. The first (and perhaps most natural)

¹⁰ Empirically, measuring social welfare in the standard framework is based on households' revealed preferences, with several restrictions that are well summarised by Slesnick (1998) – among which households are supposed to constantly maximise their utility.

interpretation was to consider them as departures from logical rules. That is, biases, which were considered as *mistakes* in the heuristics-and-biases programme, were first considered as mistakes of *logical reasoning*. There are many examples. One is the *conjunction fallacy* (Tversky and Kahneman 1974), which provides evidence that individuals often violate fundamental rules of probability theory. Another is *framing* (Tversky and Kahneman 1981), which provides evidence that individuals often violate the *invariance* principle – according to which one’s preference order between prospects should not depend on the manner in which they are described.¹¹

However, knowing that biases are various and can affect the outcome (mostly monetary, but not necessarily) of *any kind* of choice (e.g. in the health domain), mistakes could also be interpreted, more generally, as an aspect of individual psychology that makes individuals *worse off*.¹² This was the other interpretation of mistakes given by the heuristics-and-biases programme: in terms of *welfare*. From the beginning of the 1990s, the normative programme of ‘Kahneman et al.’ (Kahneman 1994; Kahneman, Wakker, and Sarin 1997; among others) was perhaps the first to use behavioural insights for welfare evaluation and policymaking – hence going beyond the interpretation of a mistake in terms of logical reasoning. Then, the influential asymmetric and libertarian paternalisms (Camerer et al. 2003; Thaler and Sunstein 2003) constituted the second major step of the growing interest of behavioural economists towards welfare evaluation. These contributions, written as manifestos, hold that ‘errors identified by behavioral research lead people not to behave in their own best interests’ (Camerer et al. 2003, p. 1211) and ‘emphasize the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower’ (Thaler and Sunstein 2003, p. 175).¹³

With the growing interest of eminent behavioural economists towards the evaluation and prescription of policy in the 1990s-2000s – who all propose different approaches to normative analysis (to be reviewed in Section 3) – the field of ‘normative’ behavioural economics or ‘behavioural’ normative economics was born.¹⁴ This ‘normative’ turn in behavioural economics brings us to a fundamental question

¹¹ Note, however, that language plays a decisive role in Kahneman and Tversky’s experiments, which makes it not so straightforward to interpret individuals’ responses as mistakes of logical reasoning. See Jullien (2016).

¹² One may check the English definition of a ‘mistake’ to consider the various meanings the word can take. According to Oxford Languages, a mistake is an ‘an act or judgement that is misguided or wrong’. This raises the obvious question, *wrong according to what?*

¹³ In the *nudging* literature – to be seen as the practical application of libertarian paternalism in various domains (e.g. health, education, environment, among many others) – the distinction between mistakes in terms of *logical reasoning* and *welfare* became, however, blurred. For example, eating healthier over time is a *moral/ethical* principle, while choosing an apple over a cake in intertemporal choice is a time-consistency principle, i.e. a *rational* principle. Yet a *nudger* would not necessarily be clear about which meaning of ‘mistake’ he/she holds (either in terms of logical reasoning or welfare, or both, or perhaps yet something else).

¹⁴ There are at least two terminologies than can be found in the literature to designate the ‘normative’ turn of behavioural economics: ‘normative behavioural economics’ – coined by Berg (2003) – and ‘behavioural normative economics’ – used for example by Dold and Schubert (2018). These terminologies should not be assimilated with ‘behavioural welfare economics’, which is a subset of (either) ‘normative behavioural economics’ or ‘behavioural normative economics’. This is because suggesting a normative approach that accounts for inconsistent choices does not necessarily imply that it is based on a *welfare* criterion. Sugden (2004, 2018) takes this non-welfare approach (to be reviewed in Section 3).

regarding the normative aspect of revealed preference theory: if individuals do not exhibit 'rational' preferences, what aspect of their behaviour should be taken into account in order to define what makes them better off?

2.2. Happiness, Well-being and Opportunity

There are at least three interpretations of *why* preference satisfaction matters in normative economics (McQuillin and Sugden 2012). The *happiness* interpretation, which characterises early hedonistic neoclassical economics, considers that *because individuals pursue happiness* (narrowly defined in terms of hedonic experience), it is good to let them satisfy their preferences. The *well-being* interpretation, which characterises ordinalist neoclassical economics, considers that *because individuals pursue their own interests* – and because it is assumed that each individual has his/her own conception of what is good for him/her and chooses in accordance with that – it is good to let them satisfy their preferences. The *opportunity* interpretation, which is perhaps commonly shared by liberal schools of thought in political economy, considers that *because individuals pursue their freedom of making their own decisions* (either interpreted as a means or a goal, e.g. exhausting the gains from trade), it is good to let them satisfy their preferences.

No matter which interpretation of preference satisfaction was held by economists before they started to seriously reconsider the normative aspect of revealed preference in the 1990s, it had no implication on the criterion to be used for normative analysis. As long as individuals exhibited 'rational' preferences, they were assumed to maximise their happiness, well-being, opportunity... or anything else that matters to them. However, once the rationality assumption breaks down, it could be the case that individuals still want to pursue what matters to them but fail to do so because of many biases. In this matter, preference satisfaction is no more conflated with rational choice in the behavioural paradigm. As a consequence, there is a need to propose alternative approaches to normative analysis that account for individuals' inconsistent choices. The main normative approaches to be reviewed in Section 3 are essentially different because of the different interpretations of preference satisfaction they hold (either in terms of happiness, well-being or opportunity). More fundamentally, this is because they are aligned with different historical traditions in normative economics. I now develop this point.

The happiness interpretation can be seen as going back to early hedonistic neoclassical economics promoted by Jevons (1871) and Edgeworth (1881), and inspired by the Benthamite (1780) pleasure-pain calculus. In this tradition, utility is considered as the subjective appreciation each individual has on consuming/choosing a good. But it goes even further, as the aim is to maximise each one's utility, following the Benthamite utilitarian premise of the greatest happiness of the greatest number. With respect to Benthamite utilitarianism, the aggregation of utilities is a mechanical tool serving an abstract policymaker whose goal is to increase the happiness of society. Based on the accumulation of empirical evidence from the 1970s that observed choices are inconsistent with the principles of rational choice, tenants of this approach hold that individuals cannot be assumed to maximise their happiness simply because they may fail to do so. Instead, according to this approach, we may have to propose alternative measures of happiness that are not related to observed choice but

to *hedonic experience*.¹⁵ This utilitarian ‘back to Bentham’ revival (Kahneman, Wakker, and Sarin 1997) in the late twentieth century against the ordinalist school of thought comes with two responses to the arguments of its proponents addressed before. First, modern tools are available to measure happiness (typically subjective well-being reports), and this data can be collected and treated for various purposes – even if it is more costly for the empiricist and the experimentalist when compared to observed choices (and even more restrictively, to observed prices and quantities). Second, one may not exclusively be interested in demand theory but in the normative programme of comparing individual and social situations. In this sense, it may be meaningful to undertake an approach that can provide some answers to the possibility of making interpersonal comparisons of utilities.¹⁶

The well-being interpretation faces the same concern than the happiness interpretation but addresses it in a different way. If choices are inconsistent with the principles of rational choice, well-being cannot be based on the standard framework of revealed preference anymore since the large number of biases documented in the literature suggests that individuals may not always act according to their own interests. Instead, we may have to attribute ‘special’ properties to some choices so that they can provide a better measure of one’s well-being – typically that such choices are aligned with some principles of rational choice. This approach rather follows the ordinalist tradition of welfare economics, but in a way that accounts for the possibility that individuals make *mistakes*. This means that deviations from rationality may come with a cost at the individual level (Thaler and Sunstein 2003), as well at the social level (Camerer et al. 2003). This approach can be considered as being mainstream, as it keeps the Pareto criterion for evaluating welfare, yet with some behavioural extensions – hence the label of *behavioural* welfare economics coined by Bernheim and Rangel (2007, 2009). This approach represents individuals as a collection of multiple selves revealing different preferences, but only choices that are made with ‘careful deliberation’ or without ‘rational distortions’ count (e.g. choices made under full information, cognitive capacity and perfect self-control). In this sense, if every ‘rational self’ prefers x over y , then x is judged to be better off for the individual than y . Note that twentieth-century ordinal economists disassociated rational choice theory and the notion of happiness: they presented rationality merely as maintaining a consistent ranking of alternatives. Such a ranking is commonly interpreted as individuals’ interests (or desires or values), but not as individuals’ hedonic experiences.¹⁷ In this sense, we can consider the contemporary ‘well-being’ interpretation of preference

¹⁵ Knowing that the ‘grand hedonistic tradition’ dominated early neoclassical economics for its (roughly) first fifty years, this approach – explicitly presented as a ‘back-to-Bentham’ approach – is just a special case of such hedonistic tradition. For a subtle distinction of some interpretations of the Benthamian utilitarian principle and their implications for welfare economics, see Baujard (2010). Note also that Bentham may not always be seen as a primitive predecessor of rational choice theory and welfare economics (Danchev 2016).

¹⁶ See in particular Kahneman (1999), who discusses a consequent body of empirical studies in psychology on how human sensory experience works, which (he argues) provides arguments for the possibility to make interpersonal comparisons of utilities.

¹⁷ In particular, the distinction between the *happiness* and *well-being* interpretations of preference satisfaction is that the former defines ‘what makes individuals better off’ in terms of *hedonic experience*, while the latter does not take any position on defining ‘what makes individuals better off’ and prefers to let this information private to individuals. In other words, the well-being interpretation does not refer to something in particular as what makes individuals better off. This poses serious challenges for policy applications, which practically require to *know* what makes individuals better off in order to increase social welfare (to be discussed in Section 3).

satisfaction to inherit from the ordinalist school of thought (Pareto 1909; Robbins 1932; Hicks and Allen 1934) by extending welfare analysis with behavioural insights.

The opportunity interpretation suggests yet another alternative to make normative analysis consistent with behavioural economics. If choices are inconsistent, it proposes to disentangle the idea that it is good to satisfy individuals' preferences because it is *their* preferences (the consumer sovereignty principle) from the preference satisfaction concept. In other words, rather than assuming that the consumer sovereignty principle depends on choices that are consistent with rational choice, the approach holding this interpretation instead focuses on the institutional process that allows individuals to *enhance their opportunity to choose from*, disregarding whether their choices are aligned with the principles of rational choice (Sugden 2004, 2018). Sugden's approach is neither influenced by hedonistic neoclassical economics nor by ordinalist neoclassical economics. It is influenced by the liberal schools of thought of Buchanan (1964, 1987) and Mill (1859) in political economy.¹⁸ Sugden's influence of Buchanan (1964) is on the conception of economics as a whole, which, according to Buchanan, is not a discipline about *choice* (what Robbinsian economists would hold) but about *exchange*. In addition, Sugden takes back the critique of Buchanan (1987) towards welfare economics, according to which the problem of normative evaluation is implicitly addressed to a benevolent autocrat but not to the actual individuals whose welfare is being assessed. In this matter, he holds Buchanan's (1987, pp. 248-250) *contractarian* perspective, which is the pursuit of mutual benefit by individuals who are concerned with their own interests. This implies a drastic departure from hedonistic and ordinalist neoclassical economics, which both share the standpoint of an external third party (the standard 'social planner') whose aim is to maximise individuals' happiness or well-being.

In addition, Sugden (2018) endorses three components of the liberal tradition of Mill (1859): (i) cooperation for mutual benefit is a governing principle of social life, (ii) competitive market is a network of mutually beneficial transactions and (iii) it is for each individual to judge what counts as his or her benefit. The last component relates to Sugden's rejection of asymmetric and libertarian paternalisms (Thaler and Sunstein 2003; Camerer et al. 2003), where he takes back Mill's (1859) defence of individual liberty, which only justifies paternalism when it prevents harm to others, but not when it prevents failures of well-being. Regarding the psychological foundations of Sugden's normative theory, the author sees the psychological mechanism that could support voluntary interactions among individuals who intend mutual benefit in Smith (1759). In particular, Sugden interprets Smith's (1759) sympathy (or fellow feeling) as 'the psychological substrates of desires for mutual benefit' (Sugden 2018, p. 273).¹⁹

Table 1 below summarises each interpretation of preference satisfaction (*happiness*, *well-being* and *opportunity*), with the main contemporary approach holding the given interpretation and its respective historical influences.

¹⁸ What follows synthesises Sugden's self-acknowledgement of the influence of these authors over his own work, which is therefore also based on his *own* reading of these authors. For a critical review of Sugden (2018), see Mitrouchev (2019).

¹⁹ Commenting Sugden's approach more extensively would lead us to further specifications that are outside the scope of the present article. For the respective influences of James Buchanan, J. S. Mill and Adam Smith over Sugden's normative theory, see Sugden (2018, Ch. 1, 2, 3, 6, 11).

Table 1: Different interpretations of preference satisfaction

	Contemporary approach	Historical influences
Happiness	Kahneman, Wakker, and Sarin (1997) (among others)	Bentham (1780) Jevons (1871) Edgeworth (1881)
Well-being	Thaler and Sunstein (2003) Camerer et al. (2003) Bernheim and Rangel (2007, 2009) (among others)	Pareto (1909) Robbins (1932) Hicks and Allen (1934)
Opportunity	Sugden (2004, 2018) (among others)	Smith (1759) Mill (1859) Buchanan (1964, 1987)

Since each of the three interpretations (in terms of *happiness*, *well-being* and *opportunity*) played a significant role in the history of normative economics (including welfare economics and political economy), I will focus on the normative criteria that are based on these interpretations, although other approaches holding different interpretations of preference satisfaction have also been proposed.²⁰ In the next section, I review some of the important methodological limits associated with these normative criteria. These are either based on individual utility maximisation – and thus either take the happiness or well-being interpretation of preference satisfaction (*experienced utility*, *true preference* and *choice-basis*) – or depart from the concept of rational choice by taking the opportunity interpretation of preference satisfaction (*opportunity*).

3. Methodological Review

3.1. Experienced Utility

The idea of measuring individuals' hedonic experience 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 biases, the idea is only to take what they experience in terms of pleasure and pain as the benchmark for evaluating their situation.²¹ The ethical premise of the experienced utility criterion can then be formulated as follows.

²⁰ Bhatt, Ogaki, and Yaguchi (2015) and Ogaki and Tanaka (2017, Ch. 11) proposed a 'virtue ethics' criterion, according to which it is good to satisfy individuals' meta-preferences for what is judged to be desirable for society. There is also the 'meaning' approach proposed by Loewenstein (1999), Karlsson, Loewenstein, and McCafferty (2004) and Dold and Stanton (2021), according to which it is good to realise individuals' expectations about living a meaningful life.

²¹ 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).

It is desirable to maximise individuals' experiences of pleasure (or to minimise individuals' experiences of pain).

The methodological limits of the experienced utility criterion are various. I shall restrict to the ones that are perhaps the most concerning.²² First, it is often argued that hedonism, when formulated as the maximisation of experienced utility, is too narrow a criterion to capture the many aspects of what makes life desirable. This point is well acknowledged by proponents of the experienced utility criterion, who argue that hedonic experience is only *one* component (among many others) of what makes life desirable.²³ This may be problematic for the scope and practical application of such a normative criterion, as (i) there is a wide range of life dimensions that is neglected by 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 old 'evil' of standard welfare economics: interpersonal comparisons of utilities. The theory of experienced utility measurement is constructed by 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, another individual experiences a guiding tour in the 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). Obviously, debating on whether cardinality is relevant for both descriptive and normative economics would lead us to a well-known and long strand in the debate between the utilitarian and ordinalist schools of thought – which is outside the scope of the present paper. The main point is that the cardinal assumption of preferences is far from being standard and may lead to important and well-known controversies (see e.g. Fleurbaey and Hammond (2004) for a review).

Third, experienced utility measurement was reconsidered by Kahneman himself in an interview given to the *Hareetz* newspaper (Kahneman 2018). The theoretical construction of experienced utility measurement is based on *moment utility*: what is experienced here and now. The main point made by Kahneman (1999) used to be 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 used to argue 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 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 based on moment utilities may not be what matters to individuals' objective happiness. As the author puts it:

²² For an extensive analysis of the axiomatic foundations of the normative theory of experienced utility and its limits, see Mitrouchev (2023).

²³ See in particular 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.

'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)

3.2. True Preference

Unlike the experienced utility criterion, the true preference criterion does not give 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/lauded preferences. From the social planner's viewpoint, these preferences constitute individuals' 'normative' preferences (i.e. what they *should* prefer).²⁴ *True* preferences are defined as preferences that an individual would have had, had she not been disturbed by rational foibles, biases, errors, mistakes, anomalies, or cognitive disturbances. The representation of observed choices as a combination of true preferences and errors allows the social planner to only take true preferences as normatively relevant, i.e. as what makes individuals better off. The social planner's goal is to identify these errors and then to reconstruct/recover individuals' true preferences.²⁵

One advantage of this normative criterion is that it does not require one to interpret well-being as narrowly as the experienced utility criterion does. In this manner, it may capture different aspects of life that individuals may find valuable, therefore entailing a larger domain of what makes individuals better off than the maximisation of pleasure. In this approach, it is (presumably) up to individuals to define what their own well-being is. Moreover, this normative criterion seems to be applicable to various choice situations, as the concept of true preference does not require one to elicit individuals' hedonic experience at the moment individuals choose or do something. Rather, it is assumed the social planner can have access to individuals' true preferences so that he/she can design a policy tool in order to make individuals choose according to their true preferences. The ethical premise of the true preference criterion can then be formulated as follows. *It is desirable that individuals satisfy their preferences when they are free of cognitive errors.*

²⁴ The literature proposing 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), Thaler (2018) and Sunstein (2019). Note that in the 1970s, pioneering experimental studies aimed to capture individuals' mistakes by letting individuals reconsider their choices (MacCrimmon 1968; Moskowitz 1974; Slovic and Tversky 1974; MacCrimmon and Larsson 1979). In choice under risk, there is a recent interest of measuring individuals' preferences over principles of rational choice (Benjamin, Fontana, and Kimball 2020; Nielsen and Rehbeck 2022; Breig and Feldman 2023).

²⁵ Perhaps the most eminent and influential policy application is the *nudge* literature (Thaler and Sunstein 2009; Halpern 2015). Note that before a large number of behavioural economists got interested in the true preference criterion, some authors had already given considerable support for it. The concept of true preference follows that of Harsanyi (1977, pp. 29-30) in his defence of utilitarianism. Fine (1995) aimed at distinguishing the two concepts of true preference and observed choice from a social choice perspective. From a philosophical perspective, some authors had already defended the satisfaction of self-interested 'informed', 'rational', or 'lauded' preferences as to what constitutes goodness (Gauthier 1986, Ch. 2; Arneson 1990; Goodin 1992).

Several limits are, however, associated with this normative criterion. First, contrary to the experienced utility criterion that is psychologically well based – 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.²⁶

The fact that true preference lacks 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 problem of the true preference criterion is its inability to explain in a convincing way that individuals would be better off if they satisfy preferences that are free of cognitive errors (Berg and Gigerenzer 2010). In fact, it appears that true preferences, as currently defined by most of its proponents, seem to be nothing else than preferences that are consistent with the principles of rational choice. The fact that the heuristics-and-biases *descriptive* programme is presented as a radical departure from the *homo-oeconomicus* abstraction but that its *normative* programme is (perhaps paradoxically) based on it is commonly recognised in the critical literature.²⁷

Second, if the assumption of true preference lacks psychological explanation, proponents of the true preference criterion need alternative ways for justifying its application. In particular, the social planner needs to rely on some meta-criteria about what is judged to be a better outcome than another. This is to avoid the problem that the social planner cannot *know* what individuals’ true preferences are. Rizzo and Whitman (2009) call this problem the ‘knowledge’ problem and Rebonato (2012) the ‘interpersonal intelligibility of preferences’ problem. Several meta-criteria have been proposed by proponents of the true preference criterion, but none of them seem to be satisfying because they restrict the scope of the true preference criterion to a narrow range of applications. These meta-criteria are the following.

Dominance. When one alternative strictly dominates another one either in terms of outcome or risk, it may be assumed that the former is better than the latter. For example, the social planner may assume that individuals’ true preferences are to save the maximum amount they can (e.g. they prefer more money to less when they are retired). Based on this assumption, the social planner could set the maximum amount as the default option of the savings plan. This meta-criterion is proposed by Loewenstein and Haisley (2008). The problem 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 one does not

²⁶ Further experimental research could nonetheless provide elements for justifying or rejecting the hypothesis of true preference. See in particular Benjamin, Fontana, and Kimball (2020), Nielsen and Rehbeck (2022) and Breig and Feldman (2023), who propose different experimental designs for measuring true preferences.

²⁷ 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).

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 is based on 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, on the assumption that eating healthy, not smoking and saving more are better, the choice architecture should be framed in a way that will encourage individuals to eat healthy, not smoke and save more. Proponents of the true preference criterion who support this meta-criterion 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). However, the author also argues 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 problem is that folk beliefs only allow one to say what individuals' preferences *should be*, not what they *actually are*. Strictly speaking, characterising such a meta-criterion as 'evidential' seems misleading, as one may wonder what kind of 'evidence' folk beliefs provides about what makes individuals better off.

Self-officiating (or 'as judged by themselves'). We are 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 if individuals clearly express their willingness to lose weight, stop smoking, stop procrastinating, etc., then the true preference criterion applies. For example, if overweight 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).²⁸ 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 x to y today and y to x tomorrow) then one may fairly question which of the many individuals' preferences over time has/have moral authority over the other(s) (Mitrouchev and Buonomo 2023). From a practical viewpoint, one may also argue that economists or social planners specifically want to have a normative criterion at hand when *ex-post* feedback is *not* available (Bleichrodt, Pinto, and Wakker 2001).

Clearly negative outcomes. This meta-criterion is based on 'common sense'. It states that when one outcome is clearly at the cost of individuals' interests, it appears relatively unambiguous that the true preference criterion applies (Loewenstein and

²⁸ Simply put, paternalism is here justified under individuals' *consent*. For empirical surveys about Europeans' acceptance of *nudges*, see e.g. Reisch and Sunstein (2016), Reisch, Sunstein, and Gwozdz (2017) and Sunstein, Reisch, and Kaiser (2019).

Haisley 2008). In other words, we can see this meta-criterion as a strong version of the ‘evidential view’. Addiction, bankruptcy or paying exactly the same product at a higher price (ethical considerations such as fair trade or environmental protection left apart) are examples of clearly negative outcomes. 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. This is given the name of ‘regularisation’ by Infante, Lecouteux, and Sugden (2016) and Sugden (2022).

Third, it is largely argued in the critical literature that one of the main approaches endorsing the true preference criterion – libertarian paternalism – is an oxymoron, contrary to the position held by Thaler and Sunstein (2003, 2009). In particular, when the social planner exploits individuals’ biases to help them taking the best decision, some point out the difficulty of making liberal and paternalistic values compatible, which may sometimes even lead to a trade-off between the two values.²⁹

3.3. Choice-Basis

Like the true preference criterion, the *choice-basis* criterion also takes the well-being interpretation of preference satisfaction. 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 what indicates or constitutes well-being (somehow ‘rescuing’ the consumer sovereignty principle). The choice-basis criterion goes a step further by 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 aside any ambiguity of the individuals’ reasons for their choices).³⁰ This normative criterion is rather defended by economists who are reluctant, for either epistemic or practical reasons, to assess individuals’ mental states.³¹ Indeed, economists usually take *choices*, and not something else (e.g. subjective well-being reports) as their privileged data. In this manner, they strictly conform to the ordinalist school of thought by making individual choice (or observed preference) the main normative criterion for well-being. The ethical premise of the choice-basis criterion can be formulated as follows. *It is desirable that individuals make choices that are undistorted of biases.*

²⁹ See in particular Mitchell (2005), Rizzo and Whitman (2009, 2019), Welch and Hausman (2010), Grüne-Yanoff (2012), Rebonato (2012), Hédoin (2015, 2017), Sugden (2017b) and Scoccia (2019).

³⁰ The literature proposing the choice-basis criterion as a standard of well-being includes Bernheim and Rangel (2007, 2009), Köszegi and Rabin (2007, 2008), Salant and Rubinstein (2008), Loewenstein and Ubel (2008), Bernheim (2009), Dalton and Ghosal (2012), Rubinstein and Salant (2012), Manzini and Mariotti (2014) and Bernheim (2016). There is also a related literature, which aims to extend measures of rationality deviations (Afriat 1973; Houtman and Maks 1985; Varian 1990) with behavioural foundations. This literature includes Apesteguia and Ballester (2015), Dzielwski (2020) and Ugarte (2023). Very generally, one can say that the choice-basis approach is merely bound to the old problem of recovering preferences from choices (Samuelson 1938; Mas-Colell 1977, 1978).

³¹ See e.g. Bernheim and Rangel (2008, p. 156), Manzini and Mariotti (2014, pp. 343-344) and Bernheim (2016, pp. 24-25), who advance the standard ‘ordinalist’ argument, according to which welfare economists should evaluate individuals’ situations based on individuals’ own conception of goodness (not happiness nor true preference). They also argue that choice is a far less obscure concept because there is more available data on choice than on mental states.

We see that the ethical premise of the choice-basis criterion is almost identical to the ethical premise of the true preference criterion. The subtlety is that the privileged data is here not *preference* but *choice*. That means the social planner is not required to elicit individuals' preferences but only to identify cognitive anomalies with the help of theoretical models that formally define what a mistake is. The goal of the social planner is then to take observed choices that are undistorted of biases as the normative-relevant data. The main limits associated with this normative criterion are nonetheless the following.

First, if the choice-basis criterion is based on the same assumption than the true preference criterion (i.e. that what make individuals better off are some psychological states that are free of biases), it may be questionable how it fundamentally differs from the true preference criterion. In this sense, the choice-basis criterion seems to face the same major limit of the true preference criterion: it only accounts for situations in which distortions from rationality 'obviously' make individuals worse off.

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

'choice-basis 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)

Third, unlike the other normative criteria previously reviewed that are based on individual utility maximisation (*experienced utility* and *true preference*), only proponents of the choice-basis criterion are reluctant to say something about the ethical content of the normative-relevant domain. This position is explicitly defended by Bernheim (2016), who takes the 'ethically neutral' stance of the ordinalist school of thought. As he rightfully mentions, '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 one, there seems to be no way of avoiding ethical judgements about what makes one outcome actually better than another one. In other words, one may question how a normative criterion can be 'normative' at all if it does not presuppose what makes one outcome better than another.

3.4. Opportunity

The *opportunity* criterion departs from rational choice as the normative benchmark, which is a point commonly shared by the previous normative criteria (experienced utility, true preference and choice-basis). Recall that the latter lean on a separation between what makes individuals better off and biases. By emphasising that choices which deviate from the principles of rationality 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.³² This approach considers the individual as a unit of agency, who identifies herself with her own past, present, and future actions (Sugden 2004, p. 1018). Sugden's aim is to maintain the liberal tradition of economics against libertarian paternalism, whose purpose is to combine liberal and paternalistic principles. His two main criticisms are 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 addressees of public policy. He 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 opportunity to choose. In this matter, his approach takes the opportunity interpretation of preference satisfaction. The benefits of the opportunity criterion are twofold: (i) it avoids the problematic aspects of the true preference and choice-basis criteria of determining what a decision 'free of biases' is, and (ii) it avoids saying something about what constitutes goodness by instead letting individuals be the best judge of their own good. The ethical premise of the opportunity criterion can then be formulated as follows. *It is desirable that individuals can have more opportunities to choose from rather than less.* Like the other normative criteria previously reviewed, the opportunity criterion is, however, not unproblematic from a methodological perspective.³³

First, the opportunity criterion forbids one to make comparisons between sets that are not nested. To give an 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 according to 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 any alternative may provide more opportunity than another, the opportunity criterion is silent about evaluating opportunity sets that are not nested.³⁴ The same problem 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 may constitute a challenge for policy applications, 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 is defined as the feeling of being worse off

³² The literature on the opportunity criterion includes Sugden (2003, 2004, 2007, 2008, 2010, 2017a, 2018).

³³ 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.

³⁴ This point actually refers to a complex debate in social choice theory on how to measure opportunity, and whether opportunity is measurable at all. I come back to this point further on.

by having too many alternatives to choose from.³⁵ 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.
- *Decrease of satisfaction*. Even if individuals are not paralysed by too many choices, 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 could 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 point formulated in economic 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 grow with the increase 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 based on the consumer sovereignty principle, according to which individuals are not only the best judge of their own well-being but 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.

One may also argue that the ethical premise of 'more is better' depends very much 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 insurance). 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').

The psychological phenomenon of self-constraint is characterised as the explicit willingness to have fewer alternatives than more.³⁶ Unlike choice overload, self-constraint is something that comes from the free will of individuals (i.e. it is determined by individuals themselves and not by a third party). Therefore, it perhaps 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

³⁵ The literature on choice overload includes Iyengar and Lepper (2000), Hutchinson (2005), Botti and Iyengar (2006), among others. For meta-analyses providing mixed results, see Scheibehenne, Greifeneder, and Todd (2010) and Chernev, Böckenholt, and Goodman (2015).

³⁶ The literature on self-constraint includes Elster (1979 [1998], 1983 [2016], 2000). See also Thaler (1980), who discusses situations where individuals voluntarily restrict their choices, deliberately not choosing so as to avoid psychic costs that the choices might induce.

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 reason that is 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 base normative assessments 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 . The opportunity criterion then suffers from a theoretical paradox: 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 choice rather than less. Taking the two psychological phenomena of *choice overload* and *self-constraint* together, one limit of the opportunity criterion is that it gives no normative relevance to individual *psychology* when individuals make choices.

Third, how to measure opportunity (and whether it is measurable at all) is a complex debate in social choice theory that is far from being consensual.³⁷ In a nutshell, there are at least three competing approaches in social choice theory, which I briefly present below.

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 problem 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.³⁸ For example, it may sound relatively reasonable that the opportunity set $O_1''' = \{blue\ car, red\ car, green\ car, yellow\ car, black\ car, white\ car\}$ provides less opportunity than the opportunity set $O_2''' = \{blue\ car, bicycle, train\}$, simply because the alternatives in O_2''' are more diversified than in O_1''' .

Potential preference. Another measurement of opportunity is 'the range of preference that individuals might have had in relevant circumstances' (Sugden 1998, p. 323). This approach is supported 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

³⁷ See Pattanaik and Xu (1990), Sen (1991) and Sugden (2003, 1998, 2010) for a debate.

³⁸ 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.

individuals reasonably/morally would like to choose. In other words, potential preference requires one to define what goodness objectively is – an enterprise that liberal proponents of the opportunity interpretation of preference satisfaction would like to stay away from.

Opportunity without metric (mutual advantage). Yet another approach to opportunity, endorsed by Sugden (2010), is that opportunity cannot be measured because it would require one 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 according to 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 be disappointing for some who would be reluctant to say that there is no objective characteristic associated with opportunity (such as pure quantity or diversity).

Fourth (and along with what has been discussed previously), the metaphysical interpretation of responsibility contains an implicit axiom that one is required to accept in Sugden's approach. 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). Consider also the 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 of the opportunity criterion, where responsibility only holds if individuals are already well informed and well experienced. Education plays an essential role because it is a matter of having the 'right' type of information and *how* the information is conveyed. This is, however, an aspect neglected by Sugden (2004, 2018) in his proposition of the opportunity criterion.³⁹

4. Discussion

This article proposed a literature review of the relationship between normative and behavioural economics by focusing on the relevant approaches, history, methods and limits. The historical and methodological perspectives are intimately joined. Considering the methodological debate about the interpretation of utility and the role of psychology in individual choice that occupied the history of normative economics for more than a century, it seems that the behavioural paradigm not only imposed itself in the broad discipline of normative economics, but now faces a challenge that goes beyond making normative economics consistent with evidence of rationality deviations. Based on my historical and methodological analyses, the challenge I aim to emphasise is the rather fundamental question of what expectations we should have for a normative criterion.

³⁹ For more about the lack of psychological substance of the opportunity criterion, see Schubert (2015).

Fundamentally, normative economics is about evaluating individual or social situations with the use of normative criteria, and then recommend/prescribe public policies based on such evaluation. From this perspective, the problem of finding which aspect of individual behaviour should be taken into account for defining what makes individuals better off can be seen as a problem of what *informational basis* should be relevant for evaluating individuals' situations. In other words, if every approach previously reviewed suggests various criteria to make normative assessments, it may be useful – at least from a methodological point of view – to provide an answer to the fundamental question of *what economists actually expect from a normative criterion*. By definition, a normative criterion is a rule that tells us whether one outcome is better than another. We can formulate three particular requirements, which account for three essential questions: 'better when?', 'better according to what?' and 'better how?' Answering the first question implies that one has 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 determine the boundaries of the normative relation R , i.e. what it can and cannot evaluate.⁴⁰ Answering the second question implies that one has an ethical judgement over the normative relation 'better than'. To define a normative relation R between 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 that one has a measure of this ethical content that allows one to evaluate individuals' situations, without which no evaluation would be possible.

In short, one suggestion (among possibly many others) can be to define three requirements: (i) a *practical* requirement as the ability of a normative criterion to apply to a wide range of choice situations (the *scope* of the normative-relevant domain), (ii) an *ethical* requirement as 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), and (iii) a *measurable* requirement as the ability of a normative criterion to measure individuals' situations (the *measurability* of the normative relation). Note that for the sake of the present literature review, these requirements should not be taken at face value. It is up to the economist himself/herself to set what expectation he/she has for a normative criterion. My intention is to propose *one* reading grid (among possibly many others) for comparing the normative criteria reviewed in the present survey. This can be helpful to open the discussion on what relevant requirements we expect for a normative criterion, and therefore to advance on the debate. Table 2 below summarises the methodological limits of the normative criteria reviewed in Section 3 with respect to the *practical*, *ethical* and *measurable* requirements I propose.

⁴⁰ 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).

Table 2: Normative ‘behavioural’ criteria

		Experienced utility	True preference	Choice-basis	Opportunity
Practical requirement	<i>Can it apply to a wide range of choice situations?</i>	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)
Ethical requirement	<i>Can it capture broad aspects of life that individuals may find valuable?</i>	NO (hedonism: narrow life dimension)	YES	YES	YES
	<i>Does it actually capture broad aspects of life that individuals may find valuable?</i>	N/A	NO (struggles to preserve autonomy)	NO (presumably not ethically based)	NO (choice overload and self-constraint)
	<i>Is it supposed to be psychologically based?</i>	YES	YES	YES	NO
	<i>Is it actually psychologically based?</i>	YES	NO (inner rational self-critique)	NO (inner rational self-critique and depends on non-choice data)	N/A
Measurable requirement	<i>Does it provide a measurement of individuals’ situations?</i>	YES	YES	YES	NO
	<i>Is the measurement relatively consensual?</i>	YES	YES	YES	NO (complex debate in social choice theory)

It is up to the economist himself/herself to judge whether a positive answer ('YES') corresponds to an advantage and a negative answer ('NO') corresponds to a disadvantage (or the other way around, or none of these). As for what is of the concern of this article, I hope to have provided a helpful literature review that will stimulate promising directions of research into developing and/or proposing alternative normative criteria that can face some of the limits addressed here.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review* 8(1), 67–77.

Afriat, S. N. (1973). On a system of inequalities in demand analysis: an extension of the classical method. *International Economic Review* 14(2), 460–472.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine. *Econometrica* 21(4), 503–546.

Apesteguia, J. and M. A. Ballester (2015). A measure of rationality and welfare. *Journal of Political Economy* 123(6), 1278–1310.

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

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica* 26(102), 121–127.

Baujard, A. (2010). Collective interest versus individual interest in Bentham's *felicific calculus*. Questioning welfarism and fairness. *The European Journal of the History of Economic Thought* 17(4), 607–634.

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

Benjamin, D. J., M. A. Fontana, and M. S. Kimball (2020, October). Reconsidering risk aversion. Working Paper 28007, National Bureau of Economic Research.

Bentham, J. (1780). *An Introduction to the Principles of Morals and Legislation*. Clarendon Press, Oxford.

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. (2021). In defense of behavioral welfare economics. *Journal of Economic Methodology* 28(4), 385–400.

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.

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.

Breig, Z. and P. Feldman (2023). Revealing risky mistakes through revisions. Unpublished Working Paper.

Buchanan, J. M. (1964). What should economists do? *Southern Economic Journal* 30(3), 213–222.

Buchanan, J. M. (1987). The constitution of economic policy. *The American Economic Review* 77(3), 243–250.

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.
- Danchev, S. (2016). Was Bentham a primitive rational choice theory predecessor? *The European Journal of the History of Economic Thought* 23(2), 297–322.
- 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 M. J. Rizzo (2021). The limits of opportunity-only: context-dependence and agency in behavioral welfare economics. *Journal of Economic Methodology* 28(4), 364–373.
- 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. *Erasmus Journal for Philosophy and Economics* 14(1), 1–29.
- Dziewulski, P. (2020). Just-noticeable difference as a behavioural foundation of the critical cost-efficiency index. *Journal of Economic Theory* 188, 105071.
- 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.
- Fleurbaey, M. and P. Hammond (2004). Interpersonally comparable utility. In S. Barbera, P. Hammond, and C. Seidl (Eds.), *Handbook of Utility Theory: Volume 2 Extensions*, pp. 1179–1285. Springer.

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.

Grüne-Yanoff, T. (2022). What preferences for behavioral welfare economics? *Journal of Economic Methodology* 29(2), 153–165.

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

Hands, D. (2013). Foundations of contemporary revealed preference theory. *Erkenntnis* 78, 1081–1108.

Hands, D. W. (2010). Economics, psychology and the history of consumer choice theory. *Cambridge Journal of Economics* 34(4), 633–648.

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.

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.

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.

Heukelom, F. (2014). *Behavioral Economics: A History*. Cambridge University Press.
Hicks, J. R. and R. G. D. Allen (1934). A reconsideration of the theory of value. Part I. *Economica* 1(1), 52–76.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica* 17(66), 159–174.

Houtman, M. and J. Maks (1985). Determining all maximal data subsets consistent with revealed preference. *International Economic Review* 19, 89–104.

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. 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.

Jevons, W. (1871). *The Theory of Political Economy*. London: Macmillan.

Jullien, D. (2016). All frames created equal are not identical: on the structure of Kahneman and Tversky's framing effects. *Æconomia. History, Methodology, Philosophy* 6(2), 265–291.

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 (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47(2), 263–291.

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.

Kao, Y.-F. and K. V. Velupillai (2015). Behavioural economics: classical and modern. *The European Journal of the History of Economic Thought* 22(2), 236–271.

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. *Æconomia. 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.

MacCrimmon, K. R. and S. Larsson (1979). Utility theory: Axioms versus ‘paradoxes’. In M. Allais and O. Hagen (Eds.), *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions under Uncertainty with Allais’ Rejoinder*, pp. 333–409. Dordrecht: Springer Netherlands.

MacCrimmon, K. R. (1968). Descriptive and normative implications of the decision-theory postulates. In K. Borch and J. Mossin (Eds.), *Risk and Uncertainty: Proceedings of a Conference held by the International Economic Association*, pp. 3–32. Palgrave Macmillan UK.

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.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics* 9, 587–608.

Mas-Colell, A. (1977). The recoverability of consumers’ preferences from market demand behavior. *Econometrica* 45(6), 1409–1430.

Mas-Colell, A. (1978). On revealed preference analysis. *Review of Economic Studies* 45(1), 121–131.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. Oxford University Press.

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

Mill, J. S. (1859). *On Liberty*. Broadview Press.

Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review* 99(3), 1245–1277.

Mitrouchev, I. (2019). Normative economics without the concept of preference. *CEconomia. History, Methodology, Philosophy* 9(1), 135–147.

Mitrouchev, I. (2023). From decision utility to experienced utility: back and forth. 2023 iRisk Working Paper 4.

Mitrouchev, I. and V. Buonomo (2023). Identity, ethics and behavioural welfare economics. *Economics & Philosophy*, 1–27.

Moscatti, I. (2018). *Measuring Utility: From the Marginal Revolution to Behavioral Economics*. Oxford University Press.

Moskowitz, H. (1974). Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems. *Decision Sciences* 5(2), 225–242.

Nielsen, K. and J. Rehbeck (2022, July). When choices are mistakes. *American Economic Review* 112(7), 2237–68.

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

Pareto, V. (1909). *Manuel d'Économie Politique*. V. Giard & E. Brière.

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.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature* 36(1), 11–46.

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

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* 2009(4), 905–968.

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

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. MacMillan & Co., Limited.

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.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica* 5(17), 61–71.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica* 15(60), 243–253.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley and Sons.

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.

Schubert, C. (2015). Opportunity and preference learning. *Economics and Philosophy* 31(2), 275–295.

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

Slesnick, D. T. (1998). Empirical approaches to the measurement of welfare. *Journal of Economic Literature* 36(4), 2108–2165.

Slovic, P. and A. Tversky (1974). Who accepts Savage's axiom? *Behavioral Science* 19(6), 368–373.

Slutsky, E. (1915). Sulla teoria del bilancio del consumatore. *Giornale degli Economisti e Rivista di Statistica* 51 (Anno 26)(1), 1–26.

Smith, A. (1759). *Theory of Moral Sentiments*. London: Printed for A. Millar. Sugden, R. (1998). The metric of opportunity. *Economics and Philosophy* 14(2), 307–337.

Sugden, R. (1991). Rational choice: a survey of contributions from economics and philosophy. *The Economic Journal* 101(407), 751–785.

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. (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.

Sugden, R. (2021). A response to six comments on *The Community of Advantage*. *Journal of Economic Methodology* 28(4), 419–430.

Sugden, R. (2022). Debiasing or regularisation? Two interpretations of the concept of 'true preference' in behavioural economics. *Theory and Decision* 92(3-4), 765–784.

Sunstein, C. R. (2015). Nudges, agency, and abstraction: A reply to critics. *Review of Philosophy and Psychology* 6(3), 511–529.

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.

Thoma, J. (2021). On the possibility of an anti-paternalist behavioural welfare economics. *Journal of Economic Methodology* 28(4), 350–363.

Tversky, A. and D. Kahneman (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology* 5(2), 207–232.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: heuristics and biases. *Science* 185(4157), 1124–1131.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science* 211(4481), 453–458.

Ugarte, C. (2023). Preference recoverability from inconsistent choices. Unpublished Working Paper.

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.

Varian, H. R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics* 46(1), 125–140.

Varian, H. R. (2014). *Intermediate Microeconomics: A Modern Approach* (ninth ed.). W.W. Norton & Company.

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.