A Brief History of Normativity in Behavioural Economics

Ivan Mitrouchev. Univ. Grenoble Alpes, INRAE, CNRS, Grenoble INP, GAEL, 38000 Grenoble, France. ivan.mitrouchev@inrae.fr

Malte Dold. Pomona College. Economics Department. 425 N. College Avenue. Claremont, CA 91711. malte.dold@pomona.edu

This version: February 2025

Abstract. Behavioural economics is mostly known as a paradigm of descriptive decision making. However, less is known about how behavioural economics paved its own way to normative decision making. This article traces the meaning of normativity in behavioural economics, exploring the divergent interpretations of irrational behaviour and their implications for policymaking. Our analysis begins with the heuristics-and-biases programme in the 1970s and 1980s, which focused on rational and logical decision-making principles. Normativity in this period was tied to academic debates on how to interpret empirical violations of rationality axioms observed in lab experiments. In the 1990s, the understanding of normativity evolved. Kahneman and colleagues introduced the concept of experienced utility, a welfare framework grounded in hedonism. By the 2000s, interpretations of irrational behaviour had diversified, leading to varied policy approaches. One of the most influential was nudging, which leverages individuals' cognitive biases to help them make choices that improve their welfare "as judged by themselves". We conclude by discussing the emergence of agency-centric approaches in behavioural economics, which have gained prominence in recent years as a response to the limitations of nudging.

Keywords: logic – mistake – policy – rationality – welfare

JEL codes: B21 – D90 – I31

Acknowledgements. Ivan Mitrouchev acknowledges the financial support of the FAST project (Facilitate public Action to exist from peSTicides) conducted by the Agence Nationale de la Recherche (ANR), reference 20-PCPA-0005. A preliminary draft was presented at the 2021 International Network for Economic Method (INEM) conference. We thank the audience for comments. We also thank Niels Boissonet, Jean-Sébastien Gharbi, Wade Hands, Yao Thibaut Kpegli, Jérôme Lallement and Ramzi Mabsout for helpful comments on early versions. All mistakes remain ours.

1. Introduction

> *"By the mid-1990s, behavioral economists had two primary goals. The first was empirical: finding and documenting anomalies, both in individual and firm behavior and in market prices. The second was developing theory. [...] But there was a third goal lurking in the background: could we use behavioral economics to make the world a better place? [...] The time was right to take this on."* (Thaler 2015: 307)

Behavioural economics started as a *descriptive, explanatory and predictive* enterprise. In a series of influential contributions (Tversky and Kahneman 1973, 1974, 1981, 1986; Kahneman and Tversky 1979), the heuristics-and-biases programme 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 some principles of rational choice, and (iii) propose a novel theoretical framework to describe, explain, and predict choice from the deviations of standard decision theory.[1] Although the consequences of systematic deviations from rational choice were given some attention in the early experiments of Kahneman and Tversky, the main focus of their research programme was orientated to the theory of *rational choice*: rules for decision making such as Bayesian updating, logic, and the axioms of expected utility theory. There was, however, no particular focus on the *evaluation*, *recommendation* and *prescription* of policies based on their findings. But from the 1990s, influential behavioural economists—among them Daniel Kahneman and Richard Thaler—had directed a consequent part of their research to welfare evaluation, as well as the recommendation and prescription of policy. Because most behavioural economists largely interpreted deviations of rationality as mistakes (to be discussed), they could not seriously take observed choice as the normative standard anymore.

This article proposes a brief history of normativity in behavioural economics: how it started, how it evolved with diverging interpretations of irrational behaviour, and how it led to real-world policymaking.[2] Our analysis begins with the heuristics-and-biases programme in the 1970s-1980s, which centred on the normative principles of logical decision making (Section 2). Then, the 1990s is identified as the first attempt by Kahneman and colleagues to propose a normative theory based on behavioural insights (Section 3). During the 2000s, varying interpretations of behavioural insights gave rise to distinct normative frameworks, each resulting in markedly different policy recommendations. Among the most prominent were the "preference purification" approach and the "opportunity" approach (Section 4). In the 2010s, a new wave of approaches emerged in response to critiques of nudging and its methodological foundations. Those emphasise *agency* (self-determined choice) as the normative benchmark for policy recommendations (Section 5). We then conclude (Section 6).

[1] The origins of behavioural economics can be traced back to the 1950s with Herbert Simon's (1955, 1956) contributions. However, we begin our historical analysis with the heuristics-and-biases programme, as it has become the dominant framework in contemporary behavioural economics (Heukelom 2014).

[2] For other histories, which focus on different aspects of behavioural economics, see Lecouteux (2016), Moscati (2018) and Viale (2022). See also Heukelom (2014: Ch. 4), who provides a detailed discussion of how the descriptive/prescriptive relationship in behavioural economics stabilised, and how the normative role of rational choice theory evolved through the various stages of Kahneman-and-Tversky's research.

Before proceeding, an early awareness about our use of “normativity” throughout this article should be noted. In economics, it is common to distinguish between descriptive (“is”) and normative (“ought”) models of decision making. Descriptive models aim to summarise, explain, or predict behaviour, while normative models set standards for “good” decision making. In the standard interpretation of decision theory, “goodness” is typically understood as adherence to a rationality standard that summarises welfare-improving choices. Occasionally, decision theorists and economists introduce a third category: the prescriptive. The prescriptive is a practical extension of the normative, applied to real-world agents rather than idealised, hyper-rational individuals. It addresses the following question. “How can real people—as opposed to imaginary, idealised, super-rational people without psyches—make better choices in a way that does not do violence to their deep cognitive concerns?” (Bell et al. 1988: 9). Prescriptive analyses often draw on the logical properties of normative theories and the empirical insights from descriptive studies in order to propose interventions that help real-world individuals make better choices. An example of a prescriptive intervention is the use of multiple-frame methods. For instance, if a doctor observes that patients' decisions vary depending on how medical outcomes are framed (e.g. Frame A presents mortality rates, while Frame B uses survival rates), the doctor can present both frames to the patient (see in particular Lecouteux and Mitrouchev 2024 for a philosophical defence of this approach). This approach allows the patient to consider the information from multiple perspectives, enabling them to make a more informed and balanced decision (Bell et al. 1988: 12). Since the prescriptive is essentially a subset of the normative, we do not differentiate between the two concepts in this article. However, we will highlight instances where the prescriptive perspective emerges within the approaches we discuss.

2. 1970s-1980s: Rational Choice as The Golden Standard

One of the earliest tensions between rational decision-making models and observed human behaviour can be traced back to the famous Allais (1953) paradox, which challenged the foundational assumptions of expected utility theory. Notably, Savage himself violated the *independence of irrelevant alternatives* axiom of expected utility theory when presented with Allais' hypothetical choice task—see Mongin (2019) for a detailed historical analysis. Although Savage acknowledged that the paradox exposed a conflict between his theoretical framework and his intuitive preferences, he maintained that expected utility theory was fundamentally correct as a *normative* model of decision making. He recognised that his intuitions led to inconsistent choices, but argued these were instances of human error rather than flaws in the theory itself. Because Savage considered the *independence* axiom desirable to follow in virtue of its logical structure, he ultimately chose to align with it, suggesting that when individuals, including himself, make decisions that are inconsistent with this principle, they should reconsider their choices.[3] Although behavioural economics had not yet been recognised as a distinct field, this event marked an early instance of how the discipline would later interpret deviations from decision-making rules—specifically as *errors* in judgement, or, simply put, *mistakes*.

[3] There are deeper philosophical investigations that we leave apart on whether rationally only presupposes desirability. For example, according to Parfit (1984), to be rational is not simply to have a desire for something—as Hume would argue—but to provide reason for acting in a certain way. For a reason-based theory of rational choice, see Dietrich and List (2013).

The heuristics-and-biases programme was developed primarily by Kahneman and Tversky in the 1970s. Its aim was to explore how people make decisions and judgments under uncertainty. The programme highlighted the relationship between heuristics and biases through lab experiments where participants were asked to solve cognitive choice tasks. Heuristics are mental shortcuts people use to make decisions or solve problems under various constraints—typically temporal (quick decisions), informational (complex or incomplete information), and cognitive (limited processing capacity). Through these constraints, Kahneman and Tversky's experimental results showed that people's heuristics can lead to *biases*. Those are defined as systematic and predictable judgments that arise from reliance on heuristics. This does not mean tenants of the heuristics-and-biases programme consider a bias to be an error (or mistake) in itself, but more precisely, when it deviates *from a given benchmark*. This means that a bias is interpreted as a mistake when participants fail to provide a correct answer to a problem that is subject to objective and verifiable scrutiny.[4] This was the beginning of how normativity was understood in this programme, as Thaler (2015) describes it:

> "By 'right' I do not mean right in some moral sense; instead, I mean *logically consistent*, as prescribed by the optimization model at the heart of economic reasoning, sometimes called rational choice theory." (Thaler 2015: 25—our emphasis)

This interpretation was largely shared by Kahneman and Tversky at the time, who, in their proposition of the first generation of prospect theory, provided a general note about (supposed) self-acknowledged errors of reasoning by the decision maker who violates the axioms of expected utility theory:

> "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. In these circumstances the anomalies implied by prospect theory are expected to occur." (Kahneman and Tversky 1979: 277)

It was in fact Slovic and Tversky (1974) who explicitly highlighted the foundational role of logic in establishing the desirability of rationality axioms:

> "Many decision theorists believe that the axioms of rational choice are similar to the principles of logic in the sense that no reasonable person who understands them would wish to violate them." (368)

[4] There is a fundamental disagreement between the heuristics-and-biases programme and the fast-and-frugal heuristics programme about the descriptive and normative interpretations of rationality. In short, the fast-and-frugal-heuristics programme (Todd and Gigerenzer 2012) does not consider heuristics or deviations from rationality principles as biases. As Gigerenzer (1996: 102) puts it, "biases are not biases". Instead, it holds that some heuristics yield to "good enough" decisions that depend on the environment in which those decisions are being made. In this sense, it follows Simon's (1956) satisficing criterion for judging a "good" decision—instead of the standard maximisation criterion. More fundamentally, the disagreement between the two programmes roots in conflicting interpretations about probabilities (Bayesians *versus* Frequentists). See in particular Gigerenzer (1991, 1996) and Kahneman and Tversky (1996) for a debate.

The idea that logic instructs us how we ought or ought not to think or reason is known as the *normative status of logic* (Steinberger 2017). That is, we consider it to be a bad thing to be inconsistent over logical principles, and conversely, we consider it to be a good thing to be consistent over logical principles. Because most rationality principles (e.g. *transitivity* or *independence*) are constructed on logical principles, the same applies for rationality principles.

The first troubling event concerning the interpretation of normativity within the behavioural paradigm was related to the nature of the rationality principles being discussed. Since not all rationality principles in the decision-making models under consideration were constructed on logical foundations, this marked an early pivotal moment where the criteria for deeming rationality principles desirable became ambiguous. If not grounded in logic, what determines the normative appeal of these principles? As an illustration, consider the three heuristics identified by Kahneman and Tversky in the 1970s as follows. The first is *availability* (Tversky and Kahneman 1973, 1974), which is the tendency to estimate the likelihood of an event based on how easily examples come to mind (for example, people typically overestimate the risk of aeroplane crashes after hearing about a crash on the news). The second is *representativeness* (Tversky and Kahneman 1974), which is about judging the probability of an event based on how similar it seems to a prototype or stereotype (for example, when people know someone to be quiet and methodical, they often think he/she is more likely to be a librarian than a salesperson). The third is *anchoring* (Tversky and Kahneman 1974), i.e. the tendency to rely heavily on the first piece of information when making decisions (for example, when people negotiate a price, the initial offer strongly influences the final deal).

Availability and *representativeness* are heuristics in uncertainty that lead to probability distortions, thereby violating Bayes' theorem. For a Bayesian, any violation of Bayes' theorem is seen as illogical and thus normatively unappealing, given the normative status of logic. *Anchoring*, however, differs from these heuristics as it is not inherently tied to uncertainty and does not result in a violation of any logical principle. This reflects a tendency among most behavioural economists to treat "rational thinking" as a general normative benchmark, without specifying which rationality principles are at stakes, and (perhaps most importantly) *in virtue of what underlying principles* it is desirable to follow those.[5] The same applies for the "framing effect", identified by Tversky and Kahneman (1981), and which later played a central role in nudging. It is defined as the tendency to make different decisions or judgments based on how information is presented, rather than the actual content of the information itself. Like *anchoring*, *framing* does not inherently relate to logic either. The only connection arises when a rationality principle is formalised—such as in expected utility theory, where the *invariance* principle requires that different representations of a choice problem yield consistent preferences. Unlike formal principles, framing relies on the specific language used and

[5] To go even further, some contemporary approaches such as libertarian paternalism conflate rational thinking with well-being. This, at first glance, does not seem at odds, since standard welfare economics has always considered the maximisation of utility (and therefore rational choice) to conflate with welfare/well-being. The issue is when a policy does not justify an intervention with respect to what "wrong reasoning" has been observed in the population, and with respect to *why* it comes at the cost of social welfare (e.g. smoking). We come back to this point in the relevant section.

psychological or social cues influencing individuals' introspection. As Tversky and Kahneman (1981) noted,

> “The frame that a decision-maker adopts is controlled partly by the formulation of the problem and partly by the norms, habits, and personal characteristics of the decision-maker.” (453)

In the heuristics-and-biases programme, framing was not considered as a heuristic, but as a bias. Some biases may be inaccurate regarding what is *objectively* true (what is subject to verifiable scrutiny). For example, the *overconfidence bias*—the tendency to overestimate one’s own abilities or the accuracy of one’s judgments—and the *confirmation bias*—the tendency to seek out or interpret information in a way that confirms pre-existing beliefs or hypotheses—fall into this category. But this is not the case for some other biases like *framing*, which are rather subject to *subjective* evaluation. Perhaps the most famous example is *loss aversion*, which is the tendency to prefer avoiding losses to acquiring equivalent gains. Here it falls to the individual to judge whether behaving according to loss aversion is actually incorrect, as there is no external objective benchmark—contrary to the rules of logic—on which loss aversion can be considered as an actual “bias”, and therefore as a mistake. In a nutshell, the ambiguity in interpreting normativity arose when it became difficult to categorise all normative principles under the same nature, as some were more closely aligned with the principles of logic than others.

Following the “empirical” trend of questioning the descriptive validity of rationality, a parallel line of research during this period undertook by Keneth MacCrimmon, Herbert Moskowitz, Paul Slovic and Amos Tversky proposed experiments whose aim was to test the *normative appeal* of the actual rationality principles which were empirically violated. In this innovative line of research, preferences over axioms were elicited in specific choice problems. Subjects could revise their choices in various risk tasks, revealing if their adjustments aligned more closely with rational principles. This line of research has seen recent updates with incentivised risk-based choice tasks, as explored by Benjamin et al. (2020), Nielsen and Rehbeck (2022), Breig and Feldman (2024), and Andersson et al. (2023) in the domain of social preferences.[6]

The first attempt was made by MacCrimmon (1968), who provided choice tasks to business executives, allowing them to reconsider their responses after exposure to arguments for and against principles like *transitivity*, *independence of irrelevant alternatives*, and *dominance*. Although some participants admitted to have made "mistakes”, the author notes that these were often attributed to laziness or difficulty with the questions. Moskowitz (1974) explored the normative appeal of the *independence of irrelevant alternatives* principle through Allais-and-Morlat-type problems, presenting tasks in word, tree, and matrix formats. Participants initially made choices, then viewed other students' responses to judge logical coherence, and could discuss choices in one treatment group.[7] After discussions, participants were more likely to align with rational principles. Those without discussion opportunities showed little change in their responses. Across frames and treatments, most participants

[6] Surprisingly, however, this line of research largely halted at the end of the 1970s and lay dormant for nearly half a century. We briefly revisit this point in the concluding remarks.

[7] Note again the normative status of logic by the experimentalist, who considers *logic*, and not something else (e.g. welfare or happiness), as the normative benchmark.

tended to follow rational principles.[8] The experiment of Slovic and Tversky (1974) regarding the *independence of irrelevant alternatives* axiom, however, showed contrasting results. In their experiment, students were given choices in Allais' and Ellsberg's problems without social or corrective feedback, but with a clear understanding of the principles. Despite these conditions, the majority of participants maintained their choices even when these conflicted with rational axioms.[9]

Given that these experiments showed mixed results in approval rates, it is not surprising to assume that the divergent opinions arose from the inherently subjective nature of "goodness," with the normative status of logic itself not exempt from being deemed undesirable. In any case, value judgments about what constitutes the "good" are unavoidable, an important point that was well emphasised by MacCrimmon (1968), who suggested a pragmatic alternative on this point.

> "A descriptive theory can be judged by its explanatory or predictive ability. It is more difficult, though, to judge a normative theory. Presumably, adopting a good normative theory will lead to 'better' results. But 'better' in what sense? The criteria must be specified and will often be part of the theory itself. One condition we might expect a good normative theory to satisfy is that it should seem reasonable to individuals with expertise in the domain of usage. Thus, we should expect a good normative theory of decision to seem reasonable to successful, practising decision-makers." (3-4)

MacCrimmon (1968) recognised that, in the absence of clear criteria for what should be considered "good," it might be more appropriate to let the experts (such as the business executives in his study) determine what makes them "better off". This highlighted a trade-off in using behavioural insights for normative analysis. If the "good" is not defined *a priori* by an external observer (such as a theorist, economist, or policymaker), what is "best" in a choice task remains ambiguous, as experts may hold differing views. Conversely, if the "good" is predefined by the observer, the prescribed "best" choice could conflict with individuals' own assessments of what is best for themselves. Thus, faced with the challenge of defining a specific criterion to distinguish a good choice from a bad one, there was a need to address this gap.

3. 1990s: The First Step Towards Welfare

From the 1990s onwards, Kahneman and colleagues leaned towards the latter approach, establishing a clear-cut criterion for what is considered as the "good". In fact, it can be said that the 1990s marked a significant shift away from the logical principles of rationality, as to what constitutes *normativity*. A new research programme on happiness, led by Kahneman, set out to answer critical questions. Do people not only deviate from logical principles but also from utility maximisation when defined in traditional Benthamite terms of pleasure and pain? Could these insights be applied to

[8] The lowest rate of approval was observed among subjects in the non-discussion group after receiving feedback, where only 50% chose to align with the *independence of irrelevant alternatives* principle.

[9] Only one experimental setting showed that a majority of participants (61%) preferred to follow the axiom, while in the three other settings, a majority chose to persist in rejecting it (59%, 66%, and 80%). Slovic and Tversky (1974) specifically designed these experiments in response to concerns that MacCrimmon's (1968) study may have been influenced by experimental demand effect and due to the peculiar characteristics of the sample. In their words: "subtle pressures, in combination with the cooperativeness of subjects participating in a training course for a prestigious job, may have influenced the subjects to conform to the axioms." (369)

public policy? And if so, what conditions would be needed to build a normative theory of happiness, allowing for utility aggregation over time? The distinction between decision utility (what people choose) and experienced utility (what they feel in terms of pain/pleasure) had already been broached by Kahneman and Tversky (1984: 349-350), making the time ripe for further exploration of these ideas.

The first published experiment on this topic was conducted by Kahneman and Snell (1990), who presented evidence that individuals have difficulty accurately predicting their future experienced utility. Building on March's (1978) proposition that "decision utility" may diverge from "experienced utility," their study provided the first empirical support for that distinction. Kahneman and Varey (1991) then questioned the validity of using choice as the sole measure of utility, proposing that experienced utility comprises three elements: the experience itself, its memory, and its anticipation. Kahneman and Snell (1992) further explored whether people can predict their future hedonic experiences and concluded they often cannot. In an eight-day study where participants consumed ice cream while listening to music, they found a near-zero correlation between actual and predicted enjoyment ratings. While, according to the authors, these findings did not conclusively show an inability to predict future tastes, the authors interpreted it as *errors* in such predictions. In this sense, the "error of reasoning" perspective was directly applied to an approach where a clear-cut criterion—namely, the optimisation of hedonic states—could distinguish good choices from bad ones.

The question of establishing a normative rule for maximising hedonic states (or experienced utility) was then actively explored. Kahneman and Snell (1992) proposed three conditions for utility integration: *monotonicity* (adding pain should increase overall disutility), *non-discrimination* (two equivalent pain moments should equally contribute to overall disutility), and *additivity* (the increase in disutility from one pain point to the next should match the added experience's disutility). It was actually the first time that *axioms over welfare*—instead of axioms over *logical principles*—were proposed in the behavioural paradigm. In Kahneman et al.'s experiments, subjects endured discomforts like carrying a suitcase, sitting in a vibrating room, or standing uncomfortably. The results showed most participants violated these conditions. Notably, they found that adding pain could sometimes lower overall negative evaluation, contradicting *monotonicity*. These deviations, still viewed as errors, were then further investigated by Kahneman et al. (1993) and Fredrickson and Kahneman (1993). These studies collectively underscored that individuals often judge experiences more by their peaks and ends than by their cumulative duration, challenging traditional assumptions of utility theory and raising new questions about how we assess well-being based on memory.

Before the influential manifestos of asymmetric and libertarian paternalism (Camerer et al. 2003; Thaler and Sunstein 2003), it was in fact Kahneman (1994) who introduced the concept of paternalistic interventions based on behavioural insights, suggesting that, because individuals may fail to correctly predict their future utility, the State might possess superior knowledge about what constitutes the best outcomes for individuals compared to the individuals themselves. Kahneman's approach, however, diverged from later developments centred on paternalism. At the end of the 1990s, he aimed to construct a comprehensive theory of objective happiness rooted in subjective experiences, particularly hedonic states. Based on his experimental studies with colleagues during the 1990s, according to which (i) individuals often exhibit myopia in their decision-making processes; (ii) they may inaccurately predict their future preferences; and (iii) they tend to make flawed choices influenced by fallible memories and incorrect evaluations of past experiences, these findings prompted Kahneman (1999) to advocate for a broader definition of rationality, which included what he termed the “substantive” criterion of experienced utility. This criterion evaluates decision outcomes independently of the individuals' preference systems, marking a significant departure from traditional welfare economics that primarily focused on the satisfaction of individual preferences to assess their well-being.

A theory of experienced utility measurement was then needed, which had to establish the necessary conditions to integrate utilities across time. This endeavour is known by the seminal work of Kahneman et al. (1997). In their "back-to-Bentham" approach, they introduce a formal normative theory for what they term total experienced utility of temporally extended outcomes, encompassing a sequence of life experiences related to pleasure and pain. The authors sought to measure the temporally extended outcomes through the normative concept of "total utility," which aggregates the temporal profiles of utility that individuals experience instantaneously. They proposed that a social planner could potentially maximise the sum of total utility for each individual within an objective function. Further investigations were conducted to assess the validity of this theory (Carter and McBride 2013; Akay et al. 2023). The main focus here is not on the theory’s grounding but rather on the significant shift in *normativity* it represents. While the normative framework of the 1970s and 1980s was rooted in *logical* principles, Kahneman et al. introduced a more *ethical* perspective, establishing strict criteria for defining objective happiness.

4. 2000s: The Dominance of Libertarian Paternalism and *Nudge*

Fuelled by a series of articles on paternalism written by some of the most prominent behavioural economists of the time (Camerer et al. 2003; Thaler and Sunstein 2003), the 2000s marked a period in which the welfare implications of behavioural findings began to flourish. These articles, written as manifestos, pointed out that conventional economists opposed paternalism because they assumed individuals behave with the rationality described in neoclassical theory. However, the authors challenged this view as unrealistic, highlighting how behavioural economics demonstrates that people often “make pretty bad decisions” (Thaler and Sunstein, 2008: 5). This prompted many behavioural economists to conclude that the traditional opposition to paternalism was unjustified. Moreover, the practice of using revealed preferences as a proxy for welfare-improving choices came under scrutiny (Mitrouchev 2024). It was at this period that the challenge later labelled as the “reconciliation problem” by McQuillin and Sugden (2012) emerged. The challenge was to uphold the normative aspiration of economics—to make welfare assessments (e.g. determining that a situation A is better than a situation B)—while accounting for the findings of behavioural economics (i.e. that people often behave inconsistently with rationality principles). At the heart of this challenge lay a critical question. What alternative to preference satisfaction could serve as the normative standard guiding policy analysis? Two prominent approaches have emerged to tackle the reconciliation problem (Dold 2023; Fumagalli 2024). The first, known as the *preference purification approach*, aligns closely with the traditional framework in normative economics. The second approach, the *opportunity approach*, seeks to provide a framework for normative analysis without referring to individuals’ preferences. These two approaches stem from distinct historical traditions regarding the role of normative economics and, more broadly, the role of economics as a whole (Mitrouchev 2024). In this paper, we focus particularly on preference purification, as it is the mainstream theoretical approach on which libertarian paternalism and nudging are grounded. For a discussion of the opportunity approach, see Schubert (2015), Mitrouchev (2019), and Dold and Rizzo (2021).

The preference purification approach begins with the assumption that people's observed choices are often the result of "bad" judgments, judgments "they would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control" (Thaler and Sunstein 2008: 5). Such a statement suggests that the normative standard relies on the notion of "undistorted" or "true" preferences (Sugden 2018: 65).[10] This perspective envisions an "inner rational agent" aligned with neoclassical theory, trapped within a bias-prone "psychological shell" (Infante et al. 2016). Individuals aim to act on a core set of well-integrated preferences that remain consistent, but psychological biases often derail them during the decision-making process.[11] Favoured by many behavioural economists, this approach aligns closely with standard welfare economics, relying on the key assumption that "true" preferences are context-independent. And as Sugden (2018: 62–63) argues, without this assumption, the approach would lack a clear normative standard for evaluating revealed preferences. In practice, the logic of the preference purification approach has led behavioural economists to develop extensive lists of "biases" to explain deviations between observed choices and "true" welfare (Rizzo and Whitman 2020). Additionally, they have contributed to designing *prescriptive policies* that leverage behavioural insights with the goal to help individuals satisfy their "true" preferences. One of the most widely discussed applications of this strategy is the use of *nudge*, defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." (Thaler and Sunstein 2008: 6).

Nudges, following the logic of libertarian paternalism (Thaler and Sunstein 2003), aim to steer behaviour from "faulty" revealed preferences towards "true" preferences (the paternalistic aspect) without restricting options or significantly altering incentives (the libertarian aspect). Ultimately, they are designed to "[make] choosers better off, as judged by themselves" (Thaler and Sunstein 2008: 5). Libertarian paternalism grapples with the conflation between rational thinking—defined as "unlimited information, ability, and self-control"—and what individuals perceive as beneficial for themselves. However, a paradox arises within libertarian paternalism. The concepts of unlimited information and ability are determined by external observers, such as economists or policymakers, while the subjective assessment of what makes people better off remains inherently internal and inaccessible due to what Rizzo and Withman (2009) refer to the "knowledge problem". This issue echoes the insights of MacCrimmon

[10] As this article aims to provide a brief history of normativity in behavioural economics, we do not delve into all the nuances of the preference purification approach. Notably, we do not address generalisations, such as those proposed by Bernheim (2016, 2021) that are beyond the assumption of true preferences. For a discussion of the historical roots of the preference purification approach, see Hands (2024). For a comprehensive overview of approaches employing the preference purification approach, see Bernheim (2016) and Sugden (2018: Ch. 4).

[11] As Sunstein (2018: 7) states: "[it] is psychologically fine to think that choosers have antecedent preferences, but that because of a lack of information or a behavioural bias, their choices will not satisfy them."

(1968), who identified a critical trade-off in the realm of *normativity* within behavioural economics. On the one hand, welfare analysis can be restricted to a narrow scope, relying on decision-making experts to establish criteria for rationality (e.g. *dominance* and *transitivity*). On the other hand, if we choose to generalise the normative approach to encompass a broader population, the criteria for what constitutes a "good" decision become ambiguous and less transparent. Libertarian paternalism provides a more flexible interpretation of "mistake" due to its unclear (and potentially shifting) normative benchmark. Nudges can be based on various mechanisms, such as social norms, visual cues, or default settings (status quo), among others. Even within these categories, the normative benchmark may differ—for instance, depending on which social norms or visual cues are emphasised. Although this posed some issues in the academic literature, leading to various discussions about the philosophical and methodological problems of libertarian paternalism and nudge (Rizzo and Whitman 2009; Grüne-Yanoff 2012; Hands 2020; among many others), it did not delay the application of real-world policymaking based on the preference purification approach.[12]

5. 2010s-today: Agency-Centric Approaches

Based on the international success of *Nudge* (Thaler and Sunstein 2008), which is itself grounded on the political proposition of libertarian paternalism and the theoretical approach of preference purification, *behavioural public policy* saw remarkable growth in the 2010s. This growth was marked by a surge in policy reports spanning domains like health, transport, and finance across numerous countries, including the United States, the United Kingdom, Australia, and various European nations. The establishment of the *Behavioural Public Policy* journal in 2017 further underscored the expanding influence of behavioural insights in public policy. However, not all scholars welcomed the field's trajectory, particularly the dominance of nudging both in policy practice and academic discussion (Oliver 2023; Chater and Loewenstein 2024). Nudging focused primarily on behavioural outcomes, sidestepping deeper discussions about what it means to be fully rational and "bias-free", and whether this is genuinely a welfare-relevant state for individuals (Rizzo and Whitman 2020).

In response to the limitations of the preference purification approach, a growing number of scholars have recently advocated for *agency* as the normative yardstick in behavioural public policy.[13] In this literature, agency is understood either more objectively as the capability to form reasoned intentions and act on them (Banerjee et al. 2024) or, more subjectively, as the decision-maker's sense of competence and autonomy (Dold et al. 2024).[14] Although these approaches differ in their conceptualisation of agency, they share a common critique of strategies used in behavioural public policy that (a) treat behavioural outcomes as target variables and (b) rely on exploiting citizens' cognitive biases to achieve those outcomes. In contrast to nudges, which often capitalise on such biases, agency-centric approaches focus on improving the quality of the cognitive processes leading to choice. Proponents of

[12] See Hands (2024) for an in-depth analysis of the issues related to the preference purification approach. Alternative approaches to preference purification have also been proposed (Sugden 2004, 2018; among others), with their respective limits. see Mitrouchev (2024) for a literature review.

[13] See, for instance, Banerjee et al. (2023), Dold and Lewis (2023), Grüne-Yanoff and Hertwig (2016), Hertwig and Grüne-Yanoff (2017), Hargreaves Heap (2013, 2017, 2023).

[14] In self-determination theory, autonomy refers to "a sense of initiative and ownership in one's actions," while competence entails "the feeling of mastery, a sense that one can succeed and grow" (Ryan and Deci 2020: 1).

agency-centric approaches critique existing frameworks in behavioural public policy for defining welfare either from a first-person standpoint, which focuses solely on individuals' current preferences while neglecting the pervasive and potentially welfare-relevant influence of contextual factors on decision making, or from a third-person standpoint, which overrides individuals' revealed preferences in favour of theorists' evaluations of what constitutes welfare. In contrast, agency-centric approaches advocate for a "second-person" standpoint that prioritises individuals' capacity to engage with and reflect on the various contexts shaping their choices (Lecouteux and Mitrouchev 2024). In these approaches, there are certain conditions that are essential to evaluate welfare from the perspective of the individual himself/herself. These conditions include (1) *appropriate cognitive abilities*, (2) an *adequate range of options*, and (3) *independence from manipulation*.[15] A range of prescriptive interventions have been proposed to address these three conditions.

To address condition 1 ("cognitive abilities"), boosts have been suggested as an alternative to nudges (Grüne-Yanoff and Hertwig 2016; Hertwig and Grüne-Yanoff 2017). Boosts aim to enhance decision-makers' cognitive abilities to help them achieve their objectives "without making undue assumptions about what those objectives are" (Grüne-Yanoff & Hertwig 2016: 156). Unlike traditional consumer protection policies, which often aim to improve informational input, boosts focus on expanding agents' cognitive strategies to transform information into choices. This is achieved by training individuals to use helpful decision heuristics—simple rules of thumb that are effective in specific environments and often outperform existing cognitive strategies. One example of a boost is teaching individuals how to convert one risk format into another, such as translating relative probabilities into natural frequencies (Hertwig and Grüne-Yanoff 2017: 977). Another example is teaching individuals to use fast-and-frugal decision trees (FFTs) as a diagnostic tool, enabling quick and effective diagnoses based on only a few informational cues (Marewski and Gigerenzer 2022).[16] Unlike nudges, boosts "require the individual's active cooperation" and that "[i]ndividuals choose to engage or not to engage with a boost" (Hertwig and Grüne-Yanoff 2017: 982). For boosts to be effective, individuals must accept the training, internalise the competence, and apply it when needed. These factors are supposed to ensure that behaviour changes resulting from boosts are grounded in reason. In addition to boosts that can help individuals achieve specific objectives more effectively (*instrumental reasoning*), some proposed agency-enhancing interventions are "educational" in nature, aiming to empower individuals to reflect on their evolving, context-dependent objectives (*substantive reasoning*). To strengthen individuals' substantive reasoning capacities, agency-centric behavioural public policy "would seem naturally to be concerned with the conditions (e.g. the educational system, the media, the family, vibrancy of the arts world) that support reflection on what preferences to hold" (Hargreaves Heap 2013: 995). The effectiveness of these institutions can be evaluated by whether they enable individuals to choose life plans they fully identify with, "in the

[15] These are the three "classic" conditions of autonomy outlined by Raz (1986) in *The Morality of Freedom*. In this article, we remain neutral on whether these conditions are necessary, sufficient, or neither.

[16] The boost literature has generated an impressive array of policy proposals. These policies can be broadly categorised into three types: (1) those that enhance risk competence in scenarios where risks are known and measurable, (2) those that build domain-specific competence by teaching effective behavioural heuristics, and (3) those that train individuals to use fast-and-frugal decision trees for navigating situations of uncertainty.

sense that they have had the resources to reflect on what preferences to hold and how to act on them" (ibid.).

To address condition 2 ("adequate range of options"), agency-enhancing behavioural public policy stress the institutional foundations required for a dynamic society where individuals can explore social influences and develop through Millian "experiments in living". Such experiments expose people to a diverse range of perspectives and identities, providing the "raw material" (examples of different lifestyles and paths) to help them shape their own preferences and identities (Delmotte and Dold 2022). This perspective underscores the importance of behaviourally informed *s-frame interventions* that tackle structural barriers to such experimentation, potentially including income inequality and social obstacles to equal opportunities (Chater and Loewenstein 2023).[17] Furthermore, laws and policies that safeguard core *civil liberties* (such as freedom of expression, freedom of movement, and freedom from discrimination) can enhance agency by enabling individuals to engage in "experiments in living". Empirical evidence shows that these freedoms positively and directly impact people's sense of self-determination (Ryan and DeHaan 2023). This institutional approach sharply contrasts with the nudging paradigm, which seeks to prevent errors through *i-frame interventions*, whereas learning through "experiments in living" inherently involves making and reflecting on mistakes as part of personal growth.

To address condition 3 ("independence from manipulation") and enable individuals to pursue their own version of a flourishing life, Oliver (2018, 2022) advocates for regulatory interventions to mitigate behavioural-informed harms. Central to this proposal is the concept of "budges," a type of regulation aimed at addressing "behavioural externalities" in exchange relationships. Positioned as a middle ground between *laissez-faire* policies and overly paternalistic interventions, budges aim to ensure fairness and reduce manipulation in market transactions. Unlike nudges, which often exploit cognitive biases to steer behaviour, budges specifically target manipulative practices that undermine free and fair exchange. Oliver highlights how businesses leverage cognitive biases, such as present bias and loss aversion, to manipulate consumers through tactics like misleading advertisements or complex pricing structures. Oliver argues that these practices justify regulatory intervention by causing substantive harm in exchanges (e.g. people overconsume certain goods and show severe post-consumption regret). Examples of effective budges include regulations on payday loans or misleading gambling advertisements, both of which exploit behavioural vulnerabilities to the detriment of consumers. Ultimately, society must collectively determine what constitutes undue harm (Oliver 2023).

The approaches discussed in this section, with their emphasis on the social conditions for individual agency, offer a shift in the debate on the normative implications of behavioural economics.[18] They move away from the traditional i-frame focus on

[17] Chater and Loewenstein (2023) introduced the distinction between i-frame and s-frame analysis. I-frame analysis focuses on individual-level solutions to policy problems, assuming that adverse outcomes stem from human cognitive frailties (e.g. present bias, loss aversion). S-frame analysis emphasises systemic changes, addressing the institutional and structural factors shaping individual choices (e.g. laws, norms, narratives). In a nutshell, they aim to "fix" the rules of the game rather than the players.

[18] Sen (1999) highlights the importance of social conditions for individual agency in his influential work, *Development as Freedom*. He emphasises that "[there] is a deep complementarity between individual agency and social arrangements. It is important to give simultaneous recognition to the centrality of

modifying individual behaviour through nudges and instead highlight the importance of systemic changes (the s-frame), such as regulations and institutional reforms. While i-frame interventions often rest on problematic normative standards (as discussed earlier), they also tend to produce modest or negligible results, failing to drive meaningful societal change. By framing problems as individual shortcomings rather than structural issues, i-frame approaches risk diverting attention and resources from systemic solutions (Chater and Loewenstein 2023). Agency-centric approaches advocate for a balance between i-frame and s-frame strategies, with the latter taking precedence in addressing large-scale challenges to individual agency.

However, a critical observation is warranted. While agency-centric approaches highlight the structural conditions necessary for fostering agency, they do not definitively establish what agency is. Any normative model of agency used in public policy is inherently a “thick concept”, as it simultaneously describes and evaluates matters (Alexandrova and Fabian 2022). The effectiveness of such models should be assessed based on whether individuals feel adequately represented by them. It remains an open question whether people resonate more with the “subjective” model, emphasising the sense of agency, the “objective” model, focusing on opportunities and reasoning capabilities, or whether agency is even a significant concern for individuals. By raising awareness of how social conditions shape beliefs and preferences or prompting individuals to reflect on their own decision-making errors, agency-centric approaches might sometimes be perceived as intrusive or unsettling.

Ultimately, these are empirical issues that can be resolved by actively involving affected citizens in co-creating policies. This aligns with Chater (2022: 1), who argues that behavioural insights “do not override, but can (among many other factors) inform, our collective decision-making process. The point of behavioural insights in public policy is primarily to inform and enrich public debate when deciding the rules by which we should like to live”. Academic expertise can play a vital role in enhancing public deliberation by helping citizens and policymakers better understand the social conditions and challenges associated with individual agency. Admittedly, public deliberation is not without flaws, as it can exacerbate decision-making issues such as motivated reasoning, herd behaviour, and groupthink. Nevertheless, when guided by inclusive and well-designed rules of discourse, deliberative processes might be able to help citizens articulate and share their beliefs about agency-centric behavioural public policy (see Colin-Jaeger and Dold 2024).

6. Concluding Remarks

In this article we analysed how the concept of *normativity* evolved over time: from the advent of behavioural economics with the heuristics-and-biases programme in the 1970s, to the international success of behavioural public policy in the 2010s. Until recent years, the choice revision literature, which began in the 1970s—but was largely abandoned until it was revived in recent years—remained underutilised. Two significant gaps have emerged in this body of work. First, some authors have criticised the lack of actual evidence demonstrating that deviations from rationality principles result in individuals being worse off (Gigerenzer 2018; Sugden 2019; Rizzo and Whitman 2020). While the choice revision literature is helpful in addressing this

individual freedom and to the force of social influences on the extent and reach of individual freedom.” (Sen 1999: xii).

criticism, it remains unsatisfying because the interpretation of “mistakes” is predominantly at the discretion of the experimentalists. These experiments (Benjamin et al. 2020; Nielsen and Rehbeck 2022; Breig and Feldman 2024) indeed do not allow participants to express whether they revised their choices towards a more "rational" direction due to *errors*, or if they did so for other reasons, such as changing their minds, simply not knowing what they want, or a deliberate desire to diversify their choices. Second, there is an increasing amount of data regarding people's willingness to be nudged across various domains, supported by meta-analyses (Reisch and Sunstein 2016; among others). These findings can provide valuable insights, even though comparing experiments is challenging due to differing methodologies used to assess individuals’ willingness to be nudged.

The potential lesson from our historical analysis is that, in the absence of better empirical evidence about what genuinely improves well-being, we (as economists) should exercise caution and refrain from being overly enthusiastic about using behavioural insights to, in Thaler’s words, “make the world a better place”. Normativity became increasingly fragmented with the rise of behavioural public policy, as each policy implicitly assumes a distinct normative benchmark (e.g. exponential discounting for intertemporal decisions like savings). These benchmarks often seem to emerge “from nowhere” (Sugden 2018). Unlike earlier approaches that explicitly identified the axioms underpinning normative benchmarks, such as those isolated in the experiments of MacCrimmon (1968), the emphasis in behavioural public policy has shifted towards “common sense” policy recommendations (e.g. encouraging healthier eating) rather than clearly articulating the normative standards they are based on.

Our historical overview ended up with a discussion of recent agency-centric perspectives that prioritise the quality of individuals’ decision-making processes over presupposing “good” behavioural outcomes. While this approach holds promise, it also raises a number of complex questions. Chief among these is the challenge of conceptualising and measuring agency as a normative standard in a way that can meaningfully inform public policy analysis and institutional reform. From an external perspective, distinguishing between genuinely acting “agentically” and merely experiencing a subjective sense of agency remains inherently difficult. A subjective feeling of agency does not necessarily equate to the objective exercise of agency. Much of the current literature on agency lacks clear normative criteria for defining what constitutes a “sufficiently good” decision-making process. Even when such criteria are proposed—such as the three conditions outlined in Section 5—it remains practically challenging to determine whether an action is preceded by adequate critical judgment, accompanied by a sufficiently large choice set, and free from manipulative third-party influences. To avoid repeating some of the shortcomings of the nudge agenda, efforts to conceptualise and measure agency might need to move beyond the top-down perspective of a social planner.

REFERENCES

Akay, A., O. B. Bargain, and H. X. Jara (2023). Experienced versus decision utility: large-scale comparison for income-leisure preferences. *The Scandinavian Journal of Economics 125*(4), 823–859.

Alexandrova, A., and Fabian, M. (2022). Democratising measurement: Or why thick concepts call for coproduction. *European Journal for Philosophy of Science, 12*(1), 7.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine (The behaviour of rational man under risk: criticism of the postulates and axioms of the American school). *Econometrica 21*(4), 503–546.

Andersson, O., Lambrecht, M., & Miettinen, T. (2023). *Personal and societal conflict of distributive principles and preferences* (Helsinki GSE Discussion Paper No. 14). Helsinki Graduate School of Economics.

Banerjee, S., Grüne-Yanoff, T., John, P., and Moseley, A. (2024). It's time we put agency into Behavioural Public Policy. *Behavioural Public Policy*, 8, 789–806.

Benjamin, D. J., M. A. Fontana, and M. S Kimball (2020). Reconsidering Risk Aversion. Working Paper 28007. Working Paper Series. National Bureau of Economic Research.

Bell, D. E., Raiffa, H., and Tversky, A. (Eds.). (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge university Press.

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

Bernheim, B. D. (2021). In defense of behavioural welfare economics. *Journal of Economic Methodology*, *28*(4), 385-400.

Breig, Z., Feldman, P (2024). Revealing risky mistakes through revisions. *Journal of Risk and Uncertainty* 68, 227–254.

Busse, M. R., Pope, D. G., Pope, J. C., and Silva-Risso, J. (2015). The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, *130*(1), 371-414.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for" Asymmetric Paternalism". *University of Pennsylvania Law Review*, *151*(3), 1211-125.

Carter, S. and M. McBride (2013). Experienced utility versus decision utility: putting the 'S' in satisfaction. *The Journal of Socio-Economics 42*, 13–23.

Chater, N. (2022). What is the point of behavioral public policy? A contractarian approach. *Behavioural Public Policy*, 1-15.

Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, *46*, e147.

Colin-Jaeger, N., and Dold, M. (2024). Individual Autonomy and Public Deliberation in Behavioral Public Policy. *RG Working Paper.*

Delmotte, C., and Dold, M. (2022). Dynamic preferences and the behavioral case against sin taxes. *Constitutional Political Economy*, *33*(1), 80-99.

Dold, M. (2023). Behavioural normative economics: foundations, approaches and trends. *Fiscal Studies*, *44*(2), 137-150.

Dold, M., and Lewis, P. (2023). A neglected topos in behavioural normative economics: the opportunity and process aspect of freedom. *Behavioural Public Policy*, *7*(4), 943-953.

Dold, M., van Emmerick, E., and Fabian, M. (2024). Taking psychology seriously: a self-determination theory perspective on Robert Sugden's opportunity criterion. *Journal of Economic Methodology*, 1-18.

Dold, M. F., and Rizzo, M. J. (2021). The limits of opportunity-only: context-dependence and agency in behavioral welfare economics. *Journal of Economic Methodology, 28*(4), 364-373.

Fumagalli, R. (2024). Preferences versus opportunities: on the conceptual foundations of normative welfare economics. *Economics & Philosophy*, *40*(1), 77-101.

Gruber, J., and Koszegi, B. (2001). Is addiction 'rational'? Theory and evidence. *Quarterly Journal of Economics, 116*(4), 1261–1303.

Gruber, J., and Koszegi, B. (2002). A Theory of Government Regulation of Addictive Bads: Optimal Tax Levels and Tax Incidence for Cigarette Excise Taxation. *National Bureau of Economic Research.*

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., and Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science, 26*(1-2), 149–183.

Hands, D.W. (2020). Libertarian paternalism: taking Econs seriously. *Int Rev Econ* 67, 419–441 (2020).

Hands, D. W. (2024). On the (non) History of Preference Purification in Modern Economics. *Review of the History of Economic Thought and Methodology 1*(1), 1–42.

Hargreaves Heap, S. P. (2013). What is the meaning of behavioural economics? *Cambridge Journal of Economics*, *37*(5), 985-1000.

Hargreaves Heap, S. P. (2017). Behavioural public policy: the constitutional approach. *Behavioural Public Policy*, *1*(2), 252-265.

Hargreaves Heap, S. P. (2023). Mill's Constitution of Liberty: an alternative behavioural policy framework. *Behavioural Public Policy*, *7*(4), 933-942.

Heukelom, F. (2014). *Behavioral economics: A history*. Cambridge University Press.

Hertwig, R., and Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, *12*(6), 973-986

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.

Johnson, E. J. (2021). *The elements of choice: Why the way we decide matters*. Penguin.

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

Kahneman, D. and A. Tversky (1984). Choices, values, and frames. *American Psychol- ogist 39*(4), 341–350.

Kahneman, D. and A. Tversky (1996). On the reality of cognitive illusions. *Psychological Review 103*(3), 582–591.

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.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics 112*(2), 443–478.

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

Lecouteux, G. and I. Mitrouchev (2024). The view from *Manywhere*: normative economics with context-dependent preferences. *Economics & Philosophy 40*(2), pp. 374 – 396.

Lichtenstein, S., and P. Slovic (Eds.) (2006). *The Construction of Preference.* Cambridge University Press.

MacCrimmon, Kenneth R. (1968). Descriptive and Normative Implications of the Decision-Theory Postulates. In K. Borch and J. Mossin (Eds.) *Risk and Uncertainty*, 3–32. London: Palgrave Macmillan.

Marewski, J. N., and Gigerenzer, G. (2022). Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience*, *14*(1), 77-89.

McKenzie, C. R., Sher, S., Leong, L. M., and Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioural Economics*, *5*(3-4), 337-360.

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

Mitrouchev, I. (2024). Normative and behavioural economics: a historical and methodological review. *The European Journal of the History of Economic Thought*, *31*(4), 533–562

Mongin, P. (2019). The Allais paradox: what it became, what it really was, what it now suggests to us. *Economics & Philosophy*, *35*(3), 423–459.

Moskowitz, H. (1974). "Effects of Problem Representation and Feedback on Rational Behavior in Allais and Morlat-Type Problems." *Decision Sciences* 5 (2): 225–42.

Nielsen, K., and J. Rehbeck (2022). When Choices Are Mistakes. *American Economic Review* 112 (7): 2237–68.

Oliver, A. (2018). Nudges, shoves and budges: Behavioural economic policy frameworks. *The International journal of health planning and management*, *33*(1), 272-275.

Oliver, A. (2022). Curtailing freedoms to protect freedom: regulating against behavioural-informed infringements on a fair exchange. *Journal of European Public Policy*, *29*(12), 1982-1993.

Oliver, A. (2023). *A political economy of behavioural public policy*. Cambridge University Press.

Raz, J. (1986). *The Morality of Freedom.* Clarendon Press.

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

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

Rizzo, M. J., and Whitman, G. (2019). *Escaping paternalism: Rationality, behavioral economics, and public policy*. Cambridge University Press.

Ryan, R., and DeHaan, C. (2023). The Social Conditions for Human Flourishing: Economic and Political Influences on Basic Psychological Needs.' In R. Ryan (ed.), *The Oxford Handbook of Self-Determination Theory*. Oxford: Oxford University Press.

Christian Schubert (2015). Opportunity and preference learning. *Economics & Philosophy*, 31, pp 275-295.

Sen, A. (1999). *Development as Freedom*. Oxford University Press.

Simon, Herbert A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69 (1): 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138.

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

Steinberger, F. (2016). The normative status of logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition).

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

Sugden, R. (2018). *The community of advantage: A behavioural economist's defence of the market*. Oxford University Press.

Sugden, R. (2019). What Should Economists Do Now? In: Wagner, R. (ed.), *James M. Buchanan: A theorist of political economy and social philosophy,* 13–37. Palgrave Macmillan

Sugden, R. (2021). Normative economics without preferences. *International Review of Economics*, *68*(1), 5-19.

Sunstein, C. R. (2018). "Better off, as judged by themselves": a comment on evaluating nudges. *International Review of Economics*, *65*, 1-8.

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

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company.

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

Thaler, R., and Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth and happiness.

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.

Tversky, A. and D. Kahneman (1986). Rational choice and the framing of decisions. *The Journal of Business 59*(4), S251–S278.

Viale, R. (2022). *Nudging*. MIT Press.