

Claremont Colleges

Scholarship @ Claremont

Pomona Faculty Publications and Research

Pomona Faculty Scholarship

8-29-2023

Behavioural normative economics: foundations, approaches and trends

Malte Dold

Pomona College

Follow this and additional works at: https://scholarship.claremont.edu/pomona_fac_pub



Part of the [Economics Commons](#), [Psychology Commons](#), and the [Social Welfare Commons](#)

Recommended Citation

Dold, M. (2023), Behavioural normative economics: foundations, approaches and trends. *Fiscal Studies*, 44, 137–150. <https://doi.org/10.1111/1475-5890.12340>

This Article is brought to you for free and open access by the Pomona Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Pomona Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@claremont.edu.

Behavioural normative economics: foundations, approaches and trends

Malte Dold 

Pomona College

Correspondence

Malte Dold, Economics Department, Pomona College, Claremont, CA 91711, US.
Email: malte.dold@pomona.edu

Submitted: July 2023

Abstract

This article summarises the theoretical foundations, main approaches and current trends in the field of behavioural normative economics. It identifies *bounded rationality* and *bounded willpower* as the two core concepts that have motivated the field. Since the concepts allow for individual preferences to be context-dependent and time-inconsistent, they pose an intricate problem for standard welfare analysis. The article discusses the ways in which two prominent approaches – the preference purification approach and the opportunity approach – have tackled the problem. It argues that shortcomings in each of these approaches motivate an agency-centric perspective. The article presents two concrete policy proposals of the agency-centric approach. While this approach is promising, the article argues for pluralism in normative economics since an exclusive focus on agency can likely not do justice to the multifarious concerns that citizens hold.

KEYWORDS

agency, behavioural public policy, boosting, opportunities, true preferences, welfare

JEL CLASSIFICATION

B41, D63, D91

1 | CORE INSIGHTS OF BEHAVIOURAL ECONOMICS

By incorporating insights from psychology into economics, behavioural economics has enriched the discussions of both positive and normative economics. In positive economics, incorporating behavioural features into neoclassical models has led to better predictions about the effects of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Author. *Fiscal Studies* published by John Wiley & Sons Ltd. on behalf of Institute for Fiscal Studies.

economic incentives such as retirement savings subsidies or income tax policies.¹ In normative economics, behavioural insights have led to novel policy tools that address internalities instead of externalities, such as nudges² or sin taxes.³ Furthermore, behavioural evidence has created a lively discussion of how to conceptualise and measure individual welfare when people's revealed preferences are context-dependent and time-inconsistent.⁴ This article will focus on the insights and challenges behavioural insights have created for normative economics.

Behavioural economics starts with the assumptions of bounded rationality and bounded willpower.⁵ *Bounded rationality* captures the observation that real-world economic actors typically employ a number of decision heuristics that can lead them to deviate from the predictions of expected utility theory.⁶ Heuristics – such as anchoring, availability or representativeness – can be effective tools for making judgements under time pressure since they are cognitively efficient; yet, they can also lead to biases in judgements when they cause people to make systematic inferential errors.⁷ A core feature of heuristics is their interplay with the choice environment. When making a choice, decision heuristics cause people to look for contextual cues that help them reach a decision.⁸ This often means that what should be normatively irrelevant contextual factors can shape people's decision-making process in subtle and problematic ways. For instance, when the sun shines and it is warm outside on the day people purchase a car, they are more likely to buy a convertible.⁹ Of course, the fact that the sun is shining on the day of the purchase should not – from a rational choice perspective – play any role in the decision. Yet, in this and many other examples, people are affected by the context which limits their ability to make welfare-improving decisions.¹⁰

The second core assumption in behavioural economics, *bounded willpower*, captures the observation that, in intertemporal choice scenarios, people often discount the immediate future more heavily than the further future. Behavioural economics captures this idea in the concept of (quasi-)hyperbolic discounting.¹¹ A crucial feature of this is that it leads to *time-inconsistent preferences*: people make plans that involve small sacrifices now for a larger reward later. But when the time comes, many people do not make the small current sacrifice. They choose the smaller-sooner option (for example, they eat the chocolate cake) instead of opting for the larger-later reward (for example, abstaining to become healthy and fit).

2 | BEHAVIOURAL CHALLENGES TO STANDARD WELFARE ECONOMICS

Bounded rationality and bounded willpower create challenges for standard welfare economics. The standard approach relies on three core assumptions:¹² preferences revealed by people's choices

¹ Chetty, 2015.

² Thaler and Sunstein, 2008.

³ Gruber and Koszegi, 2001.

⁴ Dold and Schubert, 2018.

⁵ Mullainathan and Thaler, 2000. According to these authors, the third core assumption of behavioural economics is *bounded selfishness*, i.e. people hold social preferences. While the assumption has undoubtedly important implications for policymaking (see, for example, Bowles (2016)), it was arguably not as foundational for the emergence of behavioural normative economics and will therefore not be addressed in this overview article.

⁶ DellaVigna, 2009.

⁷ Kahneman, 2003.

⁸ Lichtenstein and Slovic, 2006.

⁹ Busse et al., 2015.

¹⁰ Hargreaves Heap, 2013 and 2017.

¹¹ Frederick, Loewenstein and O'Donoghue, 2002.

¹² Bernheim and Taubinsky, 2018.

TABLE 1 Dynamic inconsistency

Preference pattern	Agent (time)			Judgement
	Pre-agent (t_{PRE})	Acting-agent (t_A)	Post-agent (t_{POST})	
Consistent	O	O	O	'I did the right thing'
Local-conflict ^a	O	X	O	'I wish I had not done that'
Melancholy ^a	O	O	X	'I wish I had'
Liberation ^a	O	X	X	'I am glad I took the wrong turn'

Note: The superscript 'a' in the first column indicates cases of intrapersonal conflict. Different approaches within behavioural normative economics have proposed different ways of dealing with such cases, as we will discuss in Section 3.

Source: Reproduced from Read (2006).

are stable and coherent (*coherence assumption*), choices do not depend on inconsequential frames (*context-independence assumption*), and revealed preferences are rational in the sense that people always choose what is best for them (*rationality assumption*). These assumptions are crucial for technical reasons as they allow economists to construct well-behaved utility functions. They are also important for substantive reasons as they tightly connect choice with welfare, i.e. whatever people choose is best for them given their budget constraints. Behavioural findings challenge all three assumptions of the standard approach.

Bounded rationality pushes back against the coherence assumption. Different frames render different attributes of choices more or less salient, and thus change the weights attached to them.¹³ Consequently, choices often do not exhibit the consistency properties required to sustain the standard preference representation. This means that the context-independence assumption is also violated: seemingly irrelevant contextual factors can lead to vastly different decisions.¹⁴ Finally, context-dependence undermines the rationality assumption: people do not necessarily choose what is best for them since the power of the context can cause them to make wrong inferences, as in the convertible car purchase example above.

Similarly, bounded willpower challenges the three assumptions of standard welfare economics. Since self-control issues can lead to time-inconsistent behaviour, the coherence assumption must be questioned, at least for intertemporal choice scenarios. Bounded willpower also challenges the context-independence assumption: here, the moment in time forms the context and is often a crucial driver in the decision-making process, as we will discuss below. And finally, the rationality assumption is challenged since violations of plans can cause intention–action gaps: what people intend to do differs from what they actually do.

When preferences are context-dependent and time-inconsistent, it is hard for the economic analyst to discern which (if any) preference is the 'right' one in the sense that it is a good indicator for welfare. For example, consider a simple decision problem in which an agent has a choice between option O (saving money for larger future expenditure) and option X (spending money on smaller present consumption). There are three time periods, t_{PRE} (the period before the action), t_A (the period of the action) and t_{POST} (the period after the action). At t_{PRE} , the agent prefers to save the money for future expenditures. There are four possible preference patterns when preferences are allowed to be dynamically inconsistent: 'consistent', 'local-conflict', 'melancholy' and 'liberation' (see Table 1).

The first case is the *consistent scenario*: the pre-agent wants to save money, the acting-agent does so, and the post-agent is satisfied with the choice ('I did the right thing'). In the *local-conflict scenario*, the

¹³ Bernheim, 2021.

¹⁴ Hargreaves Heap, 2013.

pre-agent intends to save the money, but at t_A she spends it on some present consumptive purpose; after the consumption, she regrets her decision ('I wish I had not done that'). In the *melancholy scenario*, the pre-agent intends to save and the acting-agent sticks to the plan; yet, after the action, the evaluative post-agent regrets her decision ('I wish I had'). Finally, in the *liberation scenario*, the pre-agent intends to save, the acting-agent spends the money and the post-agent is ultimately happy with the decision ('I am glad I took the wrong turn'). While this is a stylised example, preference inconsistencies such as these occur frequently in many real-world scenarios that involve intertemporal trade-offs.¹⁵ The puzzle for behavioural normative economics is that it is not clear which preferences in time should be given normative priority in the welfare calculus.¹⁶ The agent's behaviour shows choice reversals: she chooses O over X under some time frames, but X over O in others. At the heart of behavioural normative economics lie problems of context-dependence and time-inconsistency of this kind.¹⁷

3 | PROMINENT APPROACHES IN BEHAVIOURAL NORMATIVE ECONOMICS

The field of behavioural normative economics has produced a number of approaches over the last two decades.¹⁸ This section will discuss two of the most prominent ones, the preference purification approach and the opportunity-based approach. Since we will see that both approaches are not unproblematic, Section 4 will present the agency-centric approach as a potential alternative.

3.1 | Preference purification approach

The preference purification approach is based on the presumption that people have true preferences from which they diverge because of decision biases or errors.¹⁹ The approach's implicit model of agency is one of an 'inner rational agent' fulfilling the axioms of neoclassical theory that is trapped in an error-inducing 'psychological shell'. In terms of modelling, this means that the approach supplements standard models with 'behavioural' elements representing the cognitive biases that arguably give rise to decision mistakes. The self-described task of the economist is to reconstruct the agent's true preferences by identifying the effects of errors in decision-making processes, and to design policies to circumvent them so that people can satisfy their true preferences.²⁰ This approach is the approach to normative analysis favoured and practised by most behavioural economists.²¹ It is also the approach most closely tied to assumptions made in standard welfare economics.^{22,23} The present article cannot do justice to all the ways behavioural economists have tried to reconstruct agents' true preferences – for example, by simulating what they would have chosen, had they not been subject to reasoning imperfections.²⁴ Yet we can explain the logic of the preference purification approach by going back to the stylised intertemporal choice problem of the previous section.

¹⁵ Chater, 2022.

¹⁶ Read, 2006.

¹⁷ Sugden, 2018.

¹⁸ For overviews, see Bernheim (2016) or Dold and Schubert (2018).

¹⁹ Sugden, 2018, 65.

²⁰ Since this is an overview article, it does not discuss all the nuances of the preference purification approach. In particular, it does not deal with generalisations (such as Bernheim (2016 and 2021) or Bernheim and Taubinsky (2018)) that dispense with the assumption of true preferences.

²¹ Sugden, 2018, 53.

²² Hargreaves Heap, 2017, 252.

²³ For instance, a crucial assumption is that true preferences are context-independent. If this were not the case, the approach would not have an unambiguous normative standard against which individuals' revealed preferences could be judged (Sugden, 2018, 62–3).

²⁴ For an overview of approaches using the preference purification method, see Bernheim (2016) or Sugden (2018, ch. 4).

We first identify a welfare-relevant domain for which the decision-maker's choices are assumed to be free from bias. In this welfare-relevant domain, preferences are true indicators of the agent's welfare, context-independent, and follow the axioms of neoclassical rationality. We then represent the true preferences with a normative objective function. In the case of intertemporal choice, this is often assumed to be exponential discounting, $U^0(u) = u^0 + \sum_{i=1}^{\infty} \delta^i u^i$.²⁵ It has the advantage that it yields rational – i.e. stable and consistent – time preferences. Then, following the logic of preference purification, we propose a psychological mechanism that intervenes between true preferences and actual choices and thus helps explain why the agent makes faulty judgements. This mechanism could be impulsivity as captured by the introduction of the β factor in the model of quasi-hyperbolic discounting: $U^0(u) = u^0 + \sum_{i=1}^{\infty} \beta \delta^i u^i$.²⁶ The only deviation from the neoclassical model is the factor β .²⁷ In the context of our example, this model can account for time inconsistency in the local-conflict scenario and the liberation scenario where the pre-agent intends to do O but the acting-agent actually does X .

At t_{PRE} , a quasi-hyperbolic pre-agent prefers the larger–later saving O over the smaller–sooner consumption reward X . This also holds for the normative model of exponential discounting that summarises the agent's true preference for all points in time as $O > X$. Yet, in the local-conflict scenario and the liberation scenario, the agent switches her preferences in the acting period t_A . Those two scenarios might be situations in which the X option is framed in a particularly attractive and salient fashion – for example, by marketing efforts of the party that wants to sell X to our agent.²⁸ The framing makes the agent behave more impulsively and go with the smaller–sooner option X instead of the larger–later reward O . In the technical sense, the framing ‘activates’ the β factor. The agent discounts the larger–later reward more heavily compared with the correct exponential discounting baseline; her revealed preference switches to $X > O$. The framing and the passing of time that activate the β factor would not have influenced a rational decision-maker; they are mere ‘ancillary conditions’.²⁹ The preference purification approach weeds out the revealed preferences $X > O$ where ancillary conditions induced an impulsivity bias that prevented the agent from implementing her truly desired course of action.

Thus far, the preference purification analysis has identified the local-conflict scenario and the liberation scenario as problematic and explained the mechanism for why the acting-agent fails to instantiate her true preferences in both cases. What does our analysis say about the post-agent's evaluative statements mentioned in the last column of Table 2? First, since $O > X$ is assumed to be the agent's true preference, the analysis does not bother with statements in the consistent scenario and the local-conflict scenario where the agent expresses verbal support for her true preferences. Yet it identifies those judgements at t_{POST} as faulty that are defending the preference relation $X > O$, i.e. the melancholy scenario (‘I wish I had’) and the liberation scenario (‘I am glad I took the wrong turn’). The quasi-hyperbolic discounting model we introduced cannot explain the psychological mechanism for why those judgements might be faulty since it is a model of prospective decision-making. Yet, to stay within the logic of the approach, behavioural economists need to identify the psychological mechanism that helps explain the source of error in retrospective judgements. For instance, in the melancholy scenario (‘I wish I had’), the agent's retrospective judgement might be explained by a

²⁵ Delmotte and Dold, 2022.

²⁶ Gruber and Koszegi's (2001 and 2002) advocacy of sin taxes is based on this approach.

²⁷ The β factor alters the discounting of the next period compared with the entire future. It reflects the individual's impulsivity concerning the present, where the relative discount rate between today and tomorrow is $\beta\delta < \delta$. However, the individual displays patience concerning the future, with a relative discount rate between future periods set at δ .

²⁸ Framing effects are widely used in marketing presumably to make people more present biased (Johnson and Ghuman, 2020). For instance, think of retail stores that diffuse subtle fragrances through heating or air-conditioning vents that make people more present biased and discount the future more heavily.

²⁹ Bernheim, 2016.

TABLE 2 Problematic cases of dynamic inconsistency according to preference purification (circled)

Preference pattern	Agent (time)			Judgement
	Pre-agent (t_{PRE})	Acting-agent (t_A)	Post-agent (t_{POST})	
Consistent	O	O	O	'I did the right thing'
Local-conflict ^a	O	X	O	'I wish I had not done that'
Melancholy ^a	O	O	X	'I wish I had'
Liberation ^a	O	X	X	'I am glad I took the wrong turn'

focusing illusion that makes her think intensely about her experienced opportunity costs.³⁰ In the evaluative period t_{POST} , the agent compares herself narrowly to a scenario where she could have spent her money in t_A on something fun. Yet, by focusing narrowly on what she has missed in *one* period, she might neglect the benefits of short-term sacrifice in many future periods. In the liberation scenario, where the agent also defends the preference relation $X > O$, the behavioural economist might identify *cognitive dissonance reduction* as the psychological mechanism that can explain the faulty judgement.

The logic of the preference purification approach has made behavioural economists end up thinking of extended lists of 'biases' to explain divergence of observed choices from true welfare.³¹ Moreover, behavioural economists have been involved in the crafting of policies that make use of behavioural insights to help people instantiate their true preferences. Some of the most widely discussed proposals that make explicit use of the preference purification strategy are nudges³² and sin taxes.³³

Nudges follow the logic of libertarian paternalism.³⁴ They are changes in the choice architecture that are supposed to steer people's behaviour away from faulty revealed preferences toward their true preference (the paternalistic part), but they do so without forbidding any options or significantly changing their economic incentives (the libertarian part). In doing so, nudges are supposed to '[make] choosers better off, as judged by themselves'.³⁵ A paradigmatic nudge is the *Save More Tomorrow* intervention,³⁶ exploiting quasi-hyperbolic discounting: people commit in advance to allocating a portion of their *future* salary increases toward saving. The trick is that the nudge does not ask people to choose between consumption *now* versus consumption *later*; rather, it asks them to choose between consumption *in the near future* and consumption *later*. In doing so, the nudge circumvents the impulsivity of the acting-agent and instantiates the (alleged) true preference for long-term saving. This article cannot do justice to the vast literature that criticises the nudging approach. Yet, in the next section, we will see that an epistemic challenge lies at the core of this approach: how can the economist-analyst know what the agent's true preferences are?

Sin taxes are another policy that instantiates the preference purification approach.³⁷ Sin taxes are excise taxes on goods and services whose consumption is tempting (such as fatty or sugary foods) but causes internalities: unaccounted-for costs that a person imposes on her future self. Similar to a Pigouvian tax which corrects market prices to reflect the social costs caused by a negative externality,

³⁰ Schkade and Kahneman, 1998.

³¹ Rizzo and Whitman, 2019.

³² Thaler and Sunstein, 2008.

³³ Gruber and Koszegi, 2001 and 2002.

³⁴ Thaler and Sunstein, 2008.

³⁵ Thaler and Sunstein, 2008, 5.

³⁶ Thaler and Benartzi, 2004.

³⁷ Delmotte and Dold, 2022.

sin taxes aim to ‘close the gap’ between prices based on revealed preferences in the case of internalities and what prices would need to be for people to follow their true preferences. As in the nudging case, the sin tax idea is not without its critics. A core challenge is that the economist-analyst needs to gain knowledge of the agent’s true preferences and the extent of her biases to individually tailor sin taxes effectively. For instance, proponents of sin taxes argue that they should be higher for individuals who have been exposed to consumption of the ‘sinful’ goods or services in the past and who are more prone to hyperbolic discounting due to their addictive effects.³⁸

3.2 | Opportunity-based approach

What if preferences are inherently unstable or inconsistent – i.e. even after we account for problems of bounded rationality and willpower, preferences might still be mutually inconsistent, context-dependent and evolving?³⁹ Or, even more fundamentally, what if preference might not be antecedent to choice but be constructed in the moment of choosing?⁴⁰ This would mean that the very existence of true preferences that lies at the heart of the preference purification approach would need to be questioned.

The opportunity-based approach to behavioural normative economics accounts for these possibilities. In fact, its main proponent, Robert Sugden, aims to propose a ‘normative economics without preferences’.⁴¹ He points out that the model of an ‘inner rational agent’ with neoclassical preferences ‘lacks psychological foundations’.⁴² So far, behavioural economics has not presented a convincing psychological mechanism that could explain ‘[the] existence of a mode of reasoning that generates preferences that satisfy the consistency axioms’.⁴³ Rather, evidence in contemporary psychology suggests that people often construct preferences ad hoc and contextually. If this is true, then the preference purification approach has a serious problem; there is a knowledge problem on behalf of the economist. She simply does not (or even cannot) know what objects are in the best interest of the choosing agent.⁴⁴ For Sugden, this means that the preference purification approach is accompanied by the latent danger that economists substitute the aim of making choosers better off, as judged by themselves, with their judgements of what they think is good for the individual.

In the context of our intertemporal choice example, Sugden’s approach would say that it is the wrong question for the economist to ask whether the agent prefers $O > X$ or $O < X$ or to try to adjudicate between the four choice scenarios (consistent, local-conflict, melancholy, liberation). There are several reasons why it is difficult to identify a preference ranking that reflects the agent’s true preferences.

First, restricting the welfare-relevant data to the subset of consistent preferences would ignore the fact that preference inconsistency does not necessarily lead to negative welfare consequences.⁴⁵ Instead, inconsistencies may well be integral to the ‘becoming’ aspect of human nature and preference learning processes.⁴⁶

Second, in the other three choice scenarios (local-conflict, melancholy, liberation), taking options off the table to eliminate inconsistencies might risk crowding out self-regulatory capacities. For, in many real-world contexts, people are aware of their preference inconsistencies and often come up with ingenious strategies to counter them.⁴⁷ An example of counteractive self-control is the website

³⁸ Gruber and Koszegi, 2002.

³⁹ Delmotte and Dold, 2022.

⁴⁰ Lichtenstein and Slovic, 2006.

⁴¹ Sugden, 2021, 5.

⁴² Sugden 2018, 13.

⁴³ Sugden, 2018, 63.

⁴⁴ Rizzo and Whitman, 2019.

⁴⁵ Arkes, Gigerenzer and Hertwig, 2016.

⁴⁶ Delmotte and Dold, 2022.

⁴⁷ Rizzo and Whitman, 2019, 251.

Stickk.com where people voluntarily sign up to donate money to a ‘bad’ cause (for example, a charity that does not align with their political views) in the case of them not sticking to their goals.

Third, one could try to evoke meta-preferences that help adjudicate between the conflicting first-order preferences. However, given the insights of behavioural economics and psychology, it may be unclear why such meta-preferences should not also be context-dependent and dynamic.⁴⁸ Moreover, meta-preferences may take on a compulsive character and become as narrow as the impulsive first-order preferences they were supposed to protect us from.⁴⁹ An example of this latter case might be a ‘workaholic’ person who has learned to overcome her first-order preferences for present consumption for the sake of long-term career plans. In doing so, the person tends to focus greatly on work and spend little time with friends and family.

Fourth, one can try to give a certain agent-in-time normative priority. However, all of them might have preferences influenced by the context they are in. For instance, the pre-agent may be overly optimistic and mispredict the costs and benefits when the action must be enacted. The acting-agent has the local knowledge of the momentary costs and benefits of an action, but neglects the long-term consequences due to present bias. Finally, giving normative priority to the post-agent is problematic since we know from psychology that evaluations are marred with various biases, such as hindsight bias, blurred memory and self-justificatory narratives (‘I am glad I took the wrong turn’). Also, the choice experience might be transformative, which makes it hard to ‘neutrally’ compare *ex-ante* and *ex-post* preferences.⁵⁰

Given these challenges, Sugden (2018, 10) says that ‘[although] there is a clear sense in which the preferences revealed by the person in the different contexts are inconsistent with one another, it is not at all obvious which (if any) of these choices is correct, or even how “correctness” should be defined’. It is unclear in which choice contexts and at which point in time people’s choices or verbal statements should be taken as normative input for nudge policymaking. As Read (2006, 681) puts it, taking sides in situations of preference inconsistencies ‘[is] more of a judgment about the “rightness” of an action, than a consideration of what it is that the person “really” wants’.

Since preferences are inherently unstable and context-dependent, Sugden suggests shifting the normative locus away from preferences to *opportunities*, where an opportunity is defined as ‘[something] that [an individual] has the power to bring about, if he so chooses’.⁵¹ Opportunities include consumption of goods and services and participation in ‘modes of life’.⁵² Sugden argues that ‘[it] is in each individual’s interest to have more opportunity rather than less’, irrespective of her preferences, the reason being that ‘to have a lot of opportunity means you can satisfy whatever preferences you might happen to have’.⁵³

This emphasis reflects his commitment to *contractarianism*.⁵⁴ According to contractarianism, the ultimate authority in evaluating an individual’s interests lies with the person herself. The individual’s values and choices are the sole basis for making normative judgements. Consequently, Sugden suggests that a contractarian principle must gain the endorsement of the individuals who are the intended recipients of economists’ recommendations.⁵⁵ He contends that the opportunity criterion meets this requirement since individuals, recognising that their preferences are influenced by context and subject to change, will understand that having more options to choose from is always beneficial to their interests. Therefore, individuals will support the opportunity criterion as the standard for making judgements in normative economics.

⁴⁸ Lewis and Dold, 2020.

⁴⁹ Elster, 2016, 21.

⁵⁰ Fabian and Dold, 2022.

⁵¹ Sugden, 2010, 49.

⁵² Sugden, 2018, 100 and 102.

⁵³ Sugden, 2018, 84.

⁵⁴ Sugden, 2018, 29–52 and 83–4.

⁵⁵ Sugden, 2018, 84.

The approach leads to a broad presumption in favour of markets that ‘[give] individuals as much opportunity as possible to do whatever they want to do’.⁵⁶ While acknowledging the necessity of government regulation, Sugden holds an overall positive view of the market due to its role as an institution that fosters voluntary exchange.⁵⁷ He believes the guiding principle for evaluating markets and institutions and shaping their design and regulation should be the extent to which they give people opportunities to act on whatever preferences they might happen to hold at any particular moment.

4 | AGENCY-CENTRIC APPROACHES: A PROMISING TREND?

Agency-centric approaches can be understood as a reaction to some of the shortcomings in the two approaches just described.⁵⁸ In this literature, agency is understood as either – more objectively – the capability to build reasoned intentions and act on them⁵⁹ or – more subjectively – as the decision-maker’s sense of competence and autonomy.^{60,61}

The agency-centric approach puts the analysis of the cognitive processes preceding choice front and centre in behavioural normative economics, whereas Sugden’s approach does not give (much) normative weight to the character of the cognitive processes preceding choice. Admittedly, the availability of a wide choice of opportunities seems important for the possibility of agency. However, opportunities alone are arguably insufficient for the capability to build reasoned intentions and the sense of competence and autonomy that would be required for people to feel they are the authors of their own lives.⁶² Agency requires also a sense of being in control of the *choice process*. A positive relationship between the size of an individual’s opportunity set and her sense of herself as an agent can easily break down.⁶³ For instance, a diminished sense of agency may result from issues such as *obfuscation* (where additional options shroud the value of the elements in the choice set), *choice overload* (where people are overwhelmed by the sheer number of options) or *decoy effects* (where the inclusion of additional items makes some items suddenly appear more attractive).⁶⁴ In such cases, it is reasonable for an individual to prefer having a smaller number of opportunities if they improve her ability to form well-reasoned intentions and act upon them.

Agency-centric approaches differ in how they conceptualise agency. Yet they are united by their critique of the preference purification approach which assumes true preferences as target outcome variables and then exploits people’s cognitive biases through nudges as means to realise those outcomes. From an agency-centric perspective, there are several problems with this approach. First, since people often do not internalise the nudged behaviour, when the intervention is taken away, the targeted outcome often disappears.⁶⁵ One possible explanation is that such interventions often fail to offer individuals compelling reasons to select the nudged option. For example, a nudge that changes the default to automatic opt-out for retirement saving plans does not help people to reflect on their

⁵⁶ Sugden, 2019, 34.

⁵⁷ Sugden, 2019, 35.

⁵⁸ Hargreaves Heap, 2013 and 2017; Grüne-Yanoff and Hertwig, 2016; Hertwig and Grüne-Yanoff, 2017; Dold and Stanton, 2021; Banerjee et al., 2023; Dold and Lewis, 2023.

⁵⁹ Banerjee et al., 2023.

⁶⁰ Dold, van Emmerick and Fabian, 2022.

⁶¹ In self-determination theory, autonomy means ‘a sense of initiative and ownership in one’s actions’, while competence comprises ‘the feeling of mastery, a sense that one can succeed and grow’ (Ryan and Deci, 2020, 1).

⁶² Dold, van Emmerick and Fabian, 2022.

⁶³ Dold and Rizzo, 2021.

⁶⁴ Crucially, the impact of menu effects tends to amplify when the choice set becomes larger. This is because individuals, faced with high search and decision costs, rely increasingly on contextual cues to reach a decision.

⁶⁵ Banerjee et al., 2023.

preferences for retirement plans. It simply exploits people's inertia. Moreover, because nudges exploit flaws in reasoning processes rather than overcoming them, they often operate 'in the dark' and neglect the possibility that the individual herself might initiate the process of behaviour change. This may contribute to backfiring effects when people become aware that they are being nudged.⁶⁶

Agency-centric policy interventions aim at *process facilitation* by enhancing people's reasoning capacities. They acknowledge that the quality of a decision process is jointly determined by (a) the quality of the information fed into the process and (b) the appropriateness of the cognitive operations that transform information into option-selections.⁶⁷ Assistive cueing would be an example for (a) and boosting would be a policy example for (b).

Assistive cues are interventions that alter the choice environment by providing high-quality information with the goal of enabling accurate choice-relevant inferences.⁶⁸ In many situations of intertemporal choice, people look for relevant cues to reach a decision since they face a choice for which they lack clear antecedent preferences or sufficient knowledge about the options, as in our intertemporal choice case X and O . Let us assume that X and O differ along various attributes. For instance, spending the money now would entail a *social component* x_S (a night out with friends), a *hedonic component* x_H (short-term relief of the stress of the day) and an *uncertainty component* x_U (possibly meeting new friends). In contrast, saving the money now would allow a bigger spend in the future; this would also entail a *social component* o_S (a vacation with one's partner), a *hedonic component* o_H (a longer break from one's busy work schedule) and an *uncertainty component* o_U (possibly meeting new friends abroad). When the agent thinks about X and O in isolation, both options sound attractive. In fact, she has a hard time putting absolute values on the attributes $\langle x_S, x_H, x_U \rangle$ and $\langle o_S, o_H, o_U \rangle$ for the two choice options X and O . Yet the situation might change when the agent does not think of each option in isolation, $\langle X \rangle$ or $\langle O \rangle$, but jointly $\langle X, O \rangle$. This will allow her to think of and evaluate each attribute category in relative terms, i.e. compare $\langle x_S, o_S \rangle$, $\langle x_H, o_H \rangle$ and $\langle x_U, o_U \rangle$. This would yield a different, and possibly richer, evaluation process as the agent does not come up with an absolute but a relative value of the elements in each category.

The experimental literature on assistive cueing lends support to the hypothesis that for options whose respective qualities differ along multiple attributes, the evaluation of one option in isolation or both options jointly leads to different estimates of their qualities. In joint evaluations, attributes that are unfamiliar typically receive a more realistic weight. In the study conducted by Hsee (1996), subjects were asked to assess either one or two candidates for a computer programming job. There were two candidates, one with the higher grade point average (GPA) and the other one with more experience. The candidate with the higher GPA, a widely known attribute, received a higher average salary from subjects when evaluated alone. In contrast, the other candidate received a higher salary in the joint evaluation scenario, honouring the value of experience for the difficult programming job. Apparently, the joint evaluation scenario provided more information and likely led to a more accurate assessment. This and other studies on joint evaluation support the idea that assistive cueing can help people make better decisions by improving the quality of the information fed into the decision-making process.⁶⁹

Boosts are a type of agency-enhancing intervention that instead aim at improving appropriateness of the cognitive operations that transform information into option-selections.⁷⁰ According to their proponents, boosts are intended to 'preserve and enable individuals' personal agency'.⁷¹ They aim at expanding the agents' competences to reach their own objectives '[without] making undue assumptions about what those objectives are'.⁷² Crucially, boosts target agents' competences by

⁶⁶ Banerjee et al., 2023.

⁶⁷ McKenzie et al., 2018.

⁶⁸ McKenzie et al., 2018.

⁶⁹ McKenzie et al., 2018.

⁷⁰ Other examples for (b) would be nudge+ or debiasing; see Banerjee et al. (2023).

⁷¹ Hertwig and Grüne-Yanoff, 2017, 982.

⁷² Grüne-Yanoff and Hertwig, 2016, 156.

training them in the use of *decision heuristics*, i.e. simplifying rules of thumb that help them achieve their goals in certain types of environments. In contrast to nudges or assistive cueing, boosts require that ‘[individuals] choose to engage or not to engage with a boost’.⁷³ Boost interventions can only be effective if the person being boosted is open and motivated to participate in the training, internalises the necessary competence and employs it in the right contexts. While these factors make boosting more resource intensive and cognitively costly than nudges, they increase the likelihood that any change in behaviour resulting from a boost is internalised and grounded in reason.

The boost literature has produced a rich set of policies.⁷⁴ We will discuss two examples of boosts that could potentially foster people’s domain-specific competence by teaching *smart heuristics* in the context of our example of intertemporal choice. One example is inspired by the study of Drexler, Fischer and Schoar (2014), who provided microentrepreneurs with training in *simple accounting heuristics* and found that these significantly improved their financial practices. The simple heuristic was ‘[a] physical rule to keep their money in two separate drawers (or purses) and to only transfer money from one drawer to the other with an explicit “IOU” note between the business and the household. At the end of the month, they could then count how much money was in the business drawer and know what their profits were’.⁷⁵ The effect of the simple heuristic was significantly larger than that of standard accounting training (such as double-entry accounting, working capital management, and investment decisions).

People can also be taught other simple accounting heuristics, such as the *envelope method*. This is a budget management heuristic that focuses on household discretionary spending – the funds remaining after essential expenses and household fixed costs. This system uses physical or digital envelopes to allocate expenses into predetermined categories. This method enables a person to gain an easy overview of how the discretionary money is spent during a month and how much is cumulatively saved toward a future spending target. Ideally, this method allows the person to get an overview of spending-and-saving patterns over time and ultimately reach a more reasoned decision between X and O in our intertemporal choice example.

The second example of a boost that could help improve intertemporal decision-making involves fostering people’s competence to alter their sense of psychological connectedness with their future selves. An agent who is aware of and can imagine being the future recipient of today’s saving is more likely to save today for future spending.⁷⁶ In accordance with this line of thinking, Hershfield et al. (2011) introduced individuals to visual representations of their future selves, generated through age-progression algorithms. Participants who engaged with their simulated future selves, presumably bridging the gap or minimising the sense of psychological disconnectedness, demonstrated a greater inclination to choose larger delayed monetary rewards over shorter immediate ones. The difference between such boosts and the nudging approach can be illustrated by referring back to the *Save More Tomorrow* (SMT) nudge. The SMT nudge does not focus on developing individuals’ competences but rather redesigns the external choice environment taking advantage of cognitive and motivational weaknesses to encourage behaviour change. By contrast, a boost would aim to enhance competences that could improve individuals’ saving habits by, for instance, strengthening the ability to psychologically connect with one’s future self and teaching simple household accounting rules.

In general, agency-enhancing interventions such as boosts are less manipulative since they address people’s reasoning capacities. They are more transparent since they require the motivational buy-in of the affected citizens, at least in the case of boosting. Considering that the interventions primarily revolve around providing information (in the case of assistive cueing) and enhancing competences (in the case of boosts), it is expected that psychological reactance towards these interventions would

⁷³ Hertwig and Grüne-Yanoff, 2017, 982.

⁷⁴ Grüne-Yanoff and Hertwig, 2016; Hertwig and Grüne-Yanoff, 2017.

⁷⁵ Drexler, Fischer and Schoar, 2014, 3.

⁷⁶ Hertwig and Grüne-Yanoff, 2017, 978.

be lower than in the case of nudging. Moreover, improving the informational stock and decision competences of decision-makers may contribute to more persistent treatment effects since people actually internalise the process of behaviour change.⁷⁷

5 | CONCLUSION

This article described two of the most prominent approaches in behavioural normative economics, the preference purification approach and the opportunity-based approach. It presented policy implications of both and analysed some of their shortcomings. Motivated by the latter, the article introduced the agency-centric approach as a potential alternative. This approach seems desirable from a methodological standpoint; by placing emphasis on the decision-making process rather than solely on behavioural outcomes, it avoids the necessity of seeking true preferences. It seems also laudable from a normative perspective since it overcomes a simple welfarist focus and puts the liberal idea of individual self-determination centre stage.

Admittedly, the agency-centric approach might create a set of new problems and challenges. Future research should explore how far the approach poses new knowledge problems for policymaking. For instance, how can policymakers know what type of policies do reliably increase individuals' sense of agency? Moreover, the approach's main policy implications – assistive cueing and boosting – still follow the logic of i-frame interventions that focus on the individual as the target of policies. Such a perspective might neglect or downplay the importance of behaviourally informed s-frame interventions that address social-structural conditions for agency, such as income inequality or social barriers to equal opportunities.⁷⁸ Finally, a crucial feature of behavioural normative economics is that policy criteria such as welfare, opportunities or agency are 'thick' concepts: they describe and evaluate at the same time.⁷⁹ If the field wants to take contractarianism seriously, its practitioners might need to study more carefully what actual people think and care about.

A potential way forward for the field might be to acknowledge the limitations of any single normative framework and seek to integrate insights from all three of the presented approaches. This would do justice to the complex and multivariate normative concerns of people. While not taking sides in the debate about which of the presented approaches is the best one, behavioural normative economics could help policymakers and affected citizens see trade-offs between the approaches and, in doing so, better understand the relative merits of welfare and other normative criteria, such as opportunity and agency.

ORCID

Malte Dold  <https://orcid.org/0000-0002-1626-9878>

REFERENCES

- Alexandrova, A. & Fabian, M. (2022), Democratising measurement: or why thick concepts call for coproduction. *European Journal for Philosophy of Science*, 12, 7.
- Arkes, H. R., Gigerenzer, G. & Hertwig, R. (2016), How bad is incoherence? *Decision*, 3, 20–39.
- Banerjee, S., Grüne-Yanoff, T., John, P. & Moseley, A. (2023), It's time we put agency into behavioural public policy. *Behavioural Public Policy*.
- 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. & Taubinsky, D. (2018), Behavioral public economics. In B. D. Bernheim, S. DellaVigna & D. Laibson (eds), *Handbook of Behavioral Economics: Foundations and Applications 1*, Elsevier.

⁷⁷ Banerjee et al., 2023.

⁷⁸ Chater and Loewenstein, 2022.

⁷⁹ Alexandrova and Fabian, 2022.

- Bowles, S. (2016), *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. Yale University Press.
- Busse, M. R., Pope, D. G., Pope, J. C. & Silva-Risso, J. (2015), The psychological effect of weather on car purchases. *Quarterly Journal of Economics*, 130(1), 371–414.
- Chater, N. (2022), What is the point of behavioural public policy? A contractarian approach. *Behavioural Public Policy*, 1–15.
- Chater, N. & Loewenstein, G. (2022), The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, in press, <https://doi.org/10.1017/S0140525X22002023>.
- Chetty, R. (2015), Behavioral economics and public policy: a pragmatic perspective. *American Economic Review*, 105(5), 1–33.
- DellaVigna, S. (2009), Psychology and economics: evidence from the field. *Journal of Economic Literature*, 47(2), 315–72.
- Delmotte, C. & Dold, M. (2022), Dynamic preferences and the behavioral case against sin taxes. *Constitutional Political Economy*, 33, 80–99.
- Dold, M. & Lewis, P. (2023), A neglected topos in behavioural normative economics: the opportunity and process aspect of freedom. *Behavioural Public Policy*, 1–11.
- Dold, M. F. & Rizzo, M. J. (2021), The limits of opportunity-only: context-dependence and agency in behavioral welfare economics. *Journal of Economic Methodology*, 28(4), 364–73.
- Dold, M. F. & Schubert, C. (2018), Toward a behavioral foundation of normative economics. *Review of Behavioral Economics*, 5(3–4), 221–41.
- Dold, M. & Stanton, A. (2021), I choose for myself, therefore I am: the contours of existentialist behavioral economics. *Erasmus Journal for Philosophy and Economics*, 14(1), 1–29.
- Dold, M., van Emmerick, E. & Fabian, M. (2022), Taking psychology seriously? A self-determination theory perspective on Sugden's opportunity criterion. <https://doi.org/10.13140/RG.2.2.12287.43685/1>.
- Drexler, A., Fischer, G. & Schoar, A. (2014), Keeping it simple: financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2), 1–31.
- Elster, J. (2016), *Sour Grapes*. Cambridge University Press.
- Fabian, M. & Dold, M. (2022), Agentic preferences: a foundation for nudging when preferences are endogenous. *Behavioural Public Policy*, 1–21.
- Frederick, S., Loewenstein, G. & O'Donoghue, T. (2002), Time discounting and time preference: a critical review. *Journal of Economic Literature*, 40(2), 351–401.
- Gruber, J. & Koszegi, B. (2001), Is addiction 'rational'? Theory and evidence. *Quarterly Journal of Economics*, 116(4), 1261–303.
- Gruber, J. & 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 (NBER), Working Paper 8777.
- Grüne-Yanoff, T. & Hertwig, R. (2016), Nudge versus boost: how coherent are policy and theory? *Minds and Machines*, 26(1), 149–83.
- 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), 252–65.
- Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L. & Bailenson, J. N. (2011), Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research*, 48, S23–37.
- Hertwig, R. & Grüne-Yanoff, T. (2017), Nudging and boosting: steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–86.
- Hsee, C. K. (1996), The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67(3), 247–57.
- Johnson, M. & Ghuman, P. (2020), *Blindsight: The (Mostly) Hidden Ways Marketing Reshapes Our Brains*. BenBella Books.
- Kahneman, D. (2003), Maps of bounded rationality: psychology for behavioral economics. *American Economic Review*, 93(5), 1449–75.
- Lewis, P. & Dold, M. (2020), James Buchanan on the nature of choice: ontology, artifactual man and the constitutional moment in political economy. *Cambridge Journal of Economics*, 44(5), 1159–79.
- Lichtenstein, S. & Slovic, P. (2006), *The Construction of Preference*. Cambridge University Press.
- McKenzie, C. R., Sher, S., Leong, L. M. & Müller-Trede, J. (2018), Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics*, 5(3–4), 337–60.
- Mullainathan, S. & Thaler, R. H. (2000), Behavioral economics. National Bureau of Economic Research (NBER), Working Paper 7948.
- Read, D. (2006), Which side are you on? The ethics of self-command. *Journal of Economic Psychology*, 27(5), 681–93.
- Rizzo, M. J. & Whitman, G. (2019), *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*. Cambridge University Press.
- Ryan, R. M. & Deci, E. L. (2020), Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860.
- Schkade, D. A. & Kahneman, D. (1998), Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9(5), 340–6.
- Sugden, R. (2010), Opportunity as mutual advantage. *Economics & Philosophy*, 26(1), 47–68.
- 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 R. Wagner (ed.), *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*, Palgrave Macmillan.
- Sugden, R. (2021), Normative economics without preferences. *International Review of Economics*, 68(1), 5–19.
- Thaler, R. & Benartzi, S. (2004), Save More Tomorrow™: using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164–87.
- Thaler, R. & Sunstein, C. (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

How to cite this article: Dold, M. (2023), Behavioural normative economics: foundations, approaches and trends. *Fiscal Studies*, 44, 137–150. <https://doi.org/10.1111/1475-5890.12340>