Behavioural Public Policy (2024), **8**, 789–806
doi:10.1017/bpp.2024.6

ARTICLE

It's time we put agency into Behavioural Public Policy

Sanchayan Banerjee[1,2], Till Grüne-Yanoff[3], Peter John[4] and Alice Moseley[5]

[1]Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [2]London School of Economics & Political Science, London, UK, [3]Royal Institute of Technology, Stockholm, Sweden, [4]King's College London, London, UK and [5]University of Exeter, Exeter, UK
Corresponding author: Sanchayan Banerjee; Email: S.Banerjee@vu.nl

(Received 23 January 2024; accepted 23 January 2024; first published online 21 March 2024)

Abstract

Promoting agency – people's ability to form intentions and to act on them freely – must become a primary objective for Behavioural Public Policy (BPP). Contemporary BPPs do not directly pursue this objective, which is problematic for many reasons. From an ethical perspective, goals like personal autonomy and individual freedom cannot be realised without nurturing citizens' agency. From an efficacy standpoint, BPPs that override agency – for example, by activating automatic psychological processes – leave citizens 'in the dark', incapable of internalising and owning the process of behaviour change. This may contribute to non-persistent treatment effects, compensatory negative spillovers or psychological reactance and backfiring effects. In this paper, we argue agency-enhancing BPPs can alleviate these ethical and efficacy limitations to longer-lasting and meaningful behaviour change. We set out philosophical arguments to help us understand and conceptualise agency. Then, we review three alternative agency-enhancing behavioural frameworks: (1) *boosts* to enhance people's competences to make better decisions; (2) *debiasing* to encourage people to reduce the tendency for automatic, impulsive responses; and (3) *nudge+* to enable citizens to think alongside nudges and evaluate them transparently. Using a multi-dimensional framework, we highlight differences in their workings, which offer comparative insights and complementarities in their use. We discuss limitations of agency-enhancing BPPs and map out future research directions.

Keywords: agency; boosts; debiasing; nudge+; BPP

Introduction

Contemporary toolkits of *Behavioural Public Policy* (BPP) often harness citizens' cognitive biases to effectuate behaviour change. For example, they use framing, defaults, or temporal repositioning to steer people in certain directions thought best for them by a third party, such as a governmental or public body. Nudged in this way, citizens might be unaware of these interventions designed to influence their behaviour. Consequently, they are left in the dark about the operational mechanisms underlying these toolkits. Even when individuals are aware, they are often influenced by these interventions already, so it is difficult for people to infer these BPPs as main drivers of their new behaviours. For example, defaulting people into a certain option causes some of them to stick with it. However, defaults typically do not provide people with the reason to choose that option: they simply endow citizens with a choice or make the choice more likely through endorsement of the nudger (Jachimowicz *et al.*, 2019), who could be biased themselves (Hallsworth *et al.*, 2018). Similarly, changing the time when a decision is made causes some people to save more; for example, the *Save More Tomorrow* intervention encouraged employees to start a pension. But it would seem odd for them to say their reason for saving more was the change in temporal distance between choice and implementation. In this way, many current BPPs do not strengthen human agency. Sometimes, they may even weaken it by harnessing causal pathways to behavioural change that cannot be (or typically are not) understood and accepted by individuals as their reasons for this behaviour. For this reason, these toolkits face ethical and effectiveness challenges, as has been highlighted recently by many scholars (John, 2018; Schmidt and Engelen, 2020; Allcott *et al.*, 2022; Ivanković and Engelen, 2022; List *et al.*, 2022; Maier *et al.*, 2022; Szaszi *et al.*, 2022). Nonetheless, given the manifold advantages of these BPPs, such as their ease of delivery and cost-effectiveness (Benartzi *et al.*, 2017), can they be modified to enhance citizens' agency to overcome their limitations whilst still retaining their practical advantages? Answering this question is what we aim to do in this paper, as we review and outline agency-enhancing interventions in BPP.

First, we present characteristic features of agency drawn from the philosophical literature. Then we set out three alternative BPP toolkits that are designed to enhance people's agency. We argue these toolkits simply ask policy-makers to commit to new procedures which enhance contemporary BPPs by making citizens more capable of owning their decisions. These agency-enhancing alternatives do not seek to displace current toolkits, like nudges or light-touch interventions, that have become commonplace in recent years. Instead, they aim to complement them, in many cases, by helping citizens build their decision-making capabilities. We discuss boosts first. Boosting improves citizens' agency through education (Grüne-Yanoff and Hertwig, 2016; 2017). It teaches individuals to make informed choices in spite of complexity, often through efficient short cuts. Then, we turn to debiasing strategies, which help people break free from their inbuilt cognitive biases and overcome negative internalities and externalities (Fischhoff, 1982; Gofen *et al.*, 2021). Finally, we review nudge+ interventions. Nudge+ is a hybrid cognitive toolkit of BPP which encourages citizens to reflect alongside nudges thereby encouraging self-ownership in the process of behaviour change (Banerjee and John, 2021; 2022; 2023a, 2023b). Our discussion of these toolkits leads us to assert that BPP should be increasingly reformed to promote agency, but this radical approach does not necessarily overrule the application of nudges. It simply suggests the adoption of a broader toolkit.

Set out this way, the spirit of this paper is first about clarification of the precise justification of agency in ethical terms and its relation to public policy decisions. We offer an account of agency enhancement based on philosophical arguments and propose how agency can be applied and embodied in practical procedures linked to BPP. Our review of these alternative BPP frameworks provides policy-makers with agency-enhancing toolkits that can promote citizens' capacities to make their own decisions. We highlight that these three approaches share in common an attempt to enhance citizens' awareness of the choice environment and their capacity to respond to it consciously. Yet they also recognise our susceptibility to framing, and the power, as well as the responsibility of the state and commercial choice architects to shape and alter the choice environment facing citizens. They do have relative advantages and disadvantages, which can be helpful in their assessment, and in working out when and where they apply. It is this relative assessment of potential disadvantages and advantages and understanding of the appropriate contexts in which agency-enhancing procedures can apply that concern the latter part of our paper.

What is agency?

Broadly speaking, agency exists wherever entities stand in a causal relationship. A causal agent, in this sense, is any entity that produces an effect. In contrast, our call to put agency into BPP employs a narrower, *normative* notion of agency connected to reasons, intentions and actions. It is the normative aspect of this agency concept that drives our critique of existing BPP practices and motivates our reform proposals.

In the philosophy of mind and action, it is widely agreed that intentionality is the mark of genuine agency. To act intentionally, in turn, means to act for a reason (Anscombe, 1957, Davidson, 1963). One can thus distinguish between intentional action (behaviour performed for a reason) and non-intentional mere behaviour (caused by something that isn't a reason). In Davidson's account, an action is a behaviour that is causally explained by a reason[1]. To illustrate, switching on the light to see what made that strange noise, is an intentional action, because the behaviour is caused by the agent's reason to better see. If the agent instead brushed against the switch, inadvertently switching on the light, she would not have acted for a reason, and her behaviour would not be intentional (we might call it 'mere behaviour').

To have a reason for action here is understood normatively: such reasons justify or make it right for someone to act in a certain way. Typically, this means that the behaviour can be rationalised by a sound practical syllogism, consisting of the agents' goals and their take on how to attain them (Audi, 1986). For example, detecting the source of the strange noise is a reason to switch on the light; flinching when one inadvertently turns on the light was brought on by a causal chain that did not involve practical reasoning at all.

Reflexes and impulses are examples of behaviour that are neither intentional nor unintentional actions. Individuals who exhibit such behaviour do not exercise agency. This distinction is important because certain BPPs might trigger such agency-less

[1] Behaviour can of course be described in multiple ways. When I switch on the light to see what made that strange noise, I act for a reason. When the light startles a burglar hiding in my house, my behaviour that caused the light to go on can be described as 'startling the burglar'. Yet my startling of the burglar was not the reason for switching on the light. To deal with this multiple description issue, Davidson's account more precisely says that an action is a behaviour such that, *under some description*, the behaviour is causally explained by a primary reason. Only if a correct causal explanation is not available for any true description of the behaviour, then the behaviour is non-intentional mere behaviour.

behaviour. To get clearer about this distinction, we draw on the 'standard theory of action', a widely accepted account of what it is to act intentionally and for reasons. It says, roughly, that something is an intentional action and done for reasons if it is caused by mental states that the agent accepts as reasons, i.e. that rationalise the action from the agent's point of view (Goldman, 1970, Mele, 2003). Such mental states in some cases might be desires and beliefs, but often will be intentions construed as irreducible mental states themselves (Bratman, 1987).

The standard theory, even in this rough presentation, provides a powerful tool for analysing the relationship between BPP and agency. Human behaviour sometimes lacks agency. This happens when reflexes (or non-rationalised) impulses cause behaviour: those reflexes and impulses cause behaviour without the agent accepting them as reasons for behaving in this way. But it also happens when subconscious or automatic causes, depending on contextual cues, are harnessed by third parties to influence behaviour. Non-consciousness or automaticity does not exclude agency: agents can recruit the relevant routines automatically through conscious intentions and plans (Clarke, 2010, Schlosser, 2019). But this causal recruitment connection can be broken when third parties manage to harness these routines to their ends[2]. Setting a default might affect people's choice, for example, but often is not (and should not) be accepted by them as the reason for their choice; the same holds for temporal changes envisaged by e.g., the *Save More Tomorrow* intervention. We claim that this happens at least sometimes with many BPP practices, especially nudging, thus generating behavioural change while at the same time deteriorating agency.

The standard theory also offers ways to mend such agency-less BPPs. What is lacking for agency in such cases is the acceptance of the behavioural causes as reasons by the agent. Note that having reasons for action, and hence acting intentionally, can come in degrees: sometimes one acts on a single reason, disregarding others that possibly speak against it. At other times, one is carefully considering all reasons for and against an action before undertaking it. Starting from a behaviour whose causes an individual accepts as reasons to a certain degree, increasing the degree to which that individual has reasons for that behaviour enhances the individual's agency with respect to that behaviour. This can be done either by (i) rationally, non-coercively convincing the individual that the present cause is an acceptable reason for that behaviour (e.g. by showing that a default option indeed has the features that would provide me with reasons for choosing it) or (ii) replacing the present cause with another one that is more acceptable as a reason for the individual.

Thus, according to our interpretation of the standard theory of agency, enhancing agency involves tightening the connection between causes and reasons for choice. In the following, we describe three different strategies that seek such enhancement, and hence 'put agency into BPP'.

[2] Agents can acquire habits and other automatic or nonconscious processes through conscious intent, e.g. by adopting certain training routines. Alternatively, agents who developed such habits or other automatic routines might endorse them *ex post,* e.g., by acknowledging that a certain habit is conducive to a goal. (Schlosser 2019, 52) The problematic cases of BPP's harnessing of habits both break the link to the earlier intentional formation as well as prevent later intentional endorsement – hence prevent intentionality and therefore deteriorate agency.

Agency-enhancing BPPs

In recent years, multiple BPP toolkits have been proposed as alternatives to standard nudge-toolkits. These devices are designed to improve citizens' abilities to reason and own their actions. In this section, we review three such strategies, namely: *boosts*, *debiasing* and *nudge+*. These toolkits fall under the broad aegis of the behavioural agency framework but are distinct in their design and goals. In other words, they share agency-enhancing characteristics but can be applied in different ways. For example, Hertwig and Grüne-Yanoff (2017) suggest a risk-literacy boost in which subjects translate one risk format into another (relative risk vs natural frequency) in order to withstand harmful framings, for example, from stakeholders with an agenda like the pharmaceutical industry. Debiasing might use the same strategy, but with another objective, namely, to improve accuracy of risk assessment. Here, nudge+ is more interested in cognition, the quality and effect of the thought process on outcomes, when thinking about a nudge which improves risk insurance decisions (say, a precision nudge).

Most of all, the causal mechanisms whereby agency is enhanced can differ to an extent, in that different procedures are involved. So boost is about the educational activities which precede the choice, and debiasing is about new procedures introduced at the time of the choice to overcome priors, biases or anchors, whereas nudge+ is about encouraging people to evaluate if a certain nudge is suitable for them, which can come before, during or sometimes even after the nudge (albeit with an option to revisit earlier choices made). Next we will summarise these different approaches to behaviour change and outline how they restore agency in citizens, before discussing the importance of agency in BPP. Then we will highlight the similarities and differences in their workings.

Boosting

Boosts aim to change behaviour by fostering people's decision-making competences in a given environment. The reason for seeking behaviour change might be paternalistic or to reduce externalities. In either case, boosts target competences rather than immediate behaviour: individuals' skills or decision tools and their match to the environment in which people make these decisions (Grüne-Yanoff and Hertwig, 2016, Hertwig and Grüne-Yanoff, 2017).

An example of a boost is training people in using simple rules of thumb for financial decision-making, without aiming to provide comprehensive accounting knowledge. One example is using a separate drawer for business and household proceeds and writing IOUs for transfers between drawers (Drexler *et al.*, 2014). Another example is training people in temptation bundling, which helps to overcome self-control problems by coupling instantly gratifying 'want' activities (e.g., watching the next episode of a habit-forming television show, checking Facebook, receiving a pedicure, eating an indulgent meal) with engagement in a 'should' behaviour that provides long-term benefits but requires the exertion of willpower (e.g., exercising at the gym, completing a paper review, spending time with a difficult relative) (Milkman *et al.*, 2013).

In summary, boosts train people in employing decision heuristics that are better for the given purposes than what they currently use. The targeted competences can be specific to a single domain (e.g., exercising) or generalise across domains (e.g., statistical literacy). A boost may enlist human cognition (e.g., decision strategies, procedural routines, motivational competences, strategic use of automatic processes), the environment (e.g., information representation or physical environment) or both. By fostering existing competences or developing new ones, boosts are designed to enable specific behaviours.

An implementation of an effective boost only offers a strategy for behavioural change. It trains people in more effective heuristics for certain types of problems, but leaves it to individual agents when to apply these heuristics. Consequently, if people endorse the objectives of a boost – say, risk-literacy, financial planning, healthy food choices, or implementing goals – they can choose to adopt it; if not, they can decline to engage with it. To this end, the boost's objective must be transparent to the boosted individual. Otherwise, they could not suppose the participation of the boosted agent and thus would inevitably fail. People can employ the 'boosted' competence to make choices for themselves. Boost interventions therefore are successful only if the individual (i) accepts training, (ii) acquires the competence and (iii) employs the competence when a problem arises. All three are reason-based choices, which ensure that boosted behaviour change is always reason-based (Grüne-Yanoff, 2018).

The application of the boosted competence to a particular situation, therefore, requires the intentional participation of the boosted agent in contrast to the nudge, which often is effective without the nudged agent's active participation. Boosts, which are by necessity reliant on the intentional participation of the boosted, therefore maintain people's agency and in many cases will even enhance it.

Debiasing

Debiasing acknowledges the role biases play in decision-making, but rather than harnessing these as a nudge does, it attempts to mitigate them; for example, by offering alternatives, encouraging reflection, or by drawing attention to the impact of biases on decision-making. Debiasing therefore disrupts instinctive decision-making, encourages reasoning on a wider set of options and builds agency, by encouraging a move from fast (type-1) to slow (type-2) information processing (Croskerry *et al.*, 2012; Brest, 2013). A variety of debiasing techniques have been explored and tested in psychology (Arkes, 1981; Larrick, 2004; Isler *et al.*, 2020). These include asking people to justify their choice in writing, inviting them to consider a range of alternative choices or opposing perspectives (i.e. 'consider the opposite'), or drawing attention to the existence of bias, just before people make, or as they make, their decision. Approaches like time delay, i.e. encouraging people to take a little longer to make their decision, have also been proposed (Byram, 1997).

In public bodies, debiasing has been advocated to reduce the likelihood of bureaucratic or political decision-making being influenced by inevitable biases, such as sunk costs, optimism or overconfidence bias, which can negatively impact on project planning, policy design, or implementation (Battaglio *et al.*, 2018; Hallsworth *et al.*, 2018; Cantarelli *et al.*, 2020; Nagtegaal *et al.*, 2020). For instance, a 'consider the opposite' strategy in project planning would encourage policy-makers to imagine a project failing in order to counter optimism bias and encourage more realistic estimates of time and costs. Debiasing has also been explored as a means of reducing ideological extremism by tackling confirmation bias (Lilienfeld *et al*., 2009), correcting misinformation (Lewandowsky *et al*., 2012) and aiding clinical judgments (Croskerry *et al*., 2012; Ludolph and Schulz, 2018).

In the context of public policy, debiasing is less common but would aim to reduce the operation of systematic cognitive biases that can lead to decision-making that may have detrimental impacts on the individual such as present bias, the affect heuristic, or framing effects, such as the anchoring effect. Two examples illustrate the potential of debiasing. The first is from the context of consumer protection, notably credit card repayment schemes. Instead of individuals being provided with one numerical anchor as their suggested minimum repayment amount, they are also asked to consider a higher repayment amount; encouraging adjustment from the initial anchor. Informing people about this possible alternative amount may help credit card customers reduce their debt more quickly (Stewart, 2009). The second example is the use of 'mixed framing' in consumer information (Godi, 2019). In the context of food labelling, whilst a public policy nudge might highlight, with a red traffic light warning, that a food contains 25% fat, a nudge by a product manufacturer will more likely emphasise that the product is 75% fat free. A mixed frame would reduce the 'misleading effects of single frames' (Godi, 2019: 2038-2039), and would indicate that the product is 75% fat free/25% fat. One can imagine further examples such as debiasing to encourage an individual to counter optimism bias in exercise routines or dietary habits (i.e., the unrealistic optimism that one will start the better diet or the exercise routine tomorrow, which often leads to inertia), which would entail highlighting the existence of this bias and encouraging them to consider an alternative strategy, i.e., to start today. These techniques are transparent to the agent and therefore, like the previous examples, have reason-based outcomes. They encourage citizen reflection to counteract powerful biases or framing effects, thus enhancing agency. In this way, there are links between debiasing and the reflective-encouraging nudge+ discussed next.

Nudge+

A nudge+, simply put, is a nudge plus reflection (Banerjee and John, 2021; 2022; 2023a, 2023b). These interventions refer to a set of hybrid policies, where citizens are encouraged to think alongside nudges. A nudge+ is a modification of two conventional behavioural policies: the nudge, which changes how choices are presented to people without necessarily banning them or making them more economically expensive (Thaler and Sunstein, 2008; 2021); and the think, which are large-scale deliberative policies, targeted at enabling citizens to reflect widely and reach an informed decision on important issues. A composite nudge-think tool, such as the nudge+, embeds mini-thinks in nudges whereby citizens are able to evaluate the nudge and its fit to their own goals and preferences before accepting it. Adding these mini-thinks to nudges allows people to minimise the undue influences that the nudge can seem to have on them.

Banerjee and John (2021; 2023a; 2023b) outline four main types of nudge+ interventions, which are defined by two factors – how and when – the reflective prompt is combined with the nudge. For example, based on how nudges are combined with the mini-thinks, they can either be one-part or two-part devices. For one-part devices, both the nudge and the think are embedded in the same tool such that they cannot be pulled apart. An example of this includes multi-pledge devices, which encourage people to break down a commitment target into smaller pledges, and then prompts them to revaluate this long-term target over time based on progress made on their short-term pledges. Say, for instance, Iris pledges to lose six stones as a new year's resolution – she is nudged into this fresh start. As the year passes, Iris might update her priorities and goals, so the nudge might become irrelevant for her over time. The one-part nudge+ brings in this flexibility. At the beginning, the multi-pledge device prompts Iris to develop a series of monthly goals, say ½ stone. At the end of every month, Iris is then informed of her progress, and asked to revaluate the long-term nudge. In this way, this nudge+ allows citizens to think about it constantly and evaluate it based on their own preferences and goals. The classic commitment device lacks this feature of citizen engagement.

For two-part devices, a mini-think is made proximate to the nudge. Banerjee *et al.* (2023) test, for example, a combination of the pledge and the default which worked as follows: in the pledge treatment, participants were asked to think whether they would like to commit to sustainable diets and if so, how would they do it; in the default treatment, they were automatically presented with sustainable options. In the nudge+ treatment, participants were first asked to evaluate if they preferred consuming sustainable diets, and based on their goals, they were spared the trouble of thinking how to do it while those who signalled a preference for sustainable diets were defaulted into them. Here, the mini-think before the default ensures that people are able to reason their own choices, so restores their agency, while the nudge ensures that people are saved from the cognitive burden of thinking throughout.

Similarly, based on the timing, the combination of this nudge and mini-think can be made simultaneous where the reflection is prompted at the time of the nudge, or it can be sequential where reflection preceded nudging or vice-versa. Taken together, we get four kinds: one- or two-part nudge+ interventions which are either simultaneous or sequential in delivery.

In a nudge+, the mini-think ensures that people are able to reason their own choices, so restores their agency, while the nudge ensures that people are saved from the cognitive burden of thinking throughout. A nudge+ is transparent and agency-enhancing by design. Thinking alongside nudging enables citizens to reevaluate their beliefs and consider their fit with the goals of the nudge. Such a process leads to perspective transformation whereby citizens end up updating their prior beliefs (in case of a belief mismatch) or reinforcing them. This rationalisation of any behavioural change, as induced by the nudge, either leads to an acceptance of the nudge ('I stuck with the default because it is in fact the best option for my purposes') or a reason-based rejection of it ('Upon reflection, I decided not to stick with the default, as the alternative option is better for my purposes'). Consequently, the nudge+ helps transform the originally non-reason-based cause of choice into a reason-based one.

The importance of agency for BPP

In this section, we discuss positive arguments in favour of building agency for citizens through BPP toolkits. In particular, we ask: why does enhancing agency matter for BPP? We outline four main reasons as to why agency should normally prevail in the design of BPP.

The first argument, following our discussion in the previous section, is that agency-enhancing toolkits improve people's abilities to decide for themselves and understand their genuine preferences. We can be persuaded to do many things and to come up with rationalisations about what we did, such as the reason we bought a brand of coffee after purchasing it (knowing full well if asked that we decided it based on prominence on the shelf or seeing someone else buy this brand). When agency is in play, there is a consideration of what we want to do so as to be consistent with other forms of preferences, life plans and value positions a person may hold. For example, if someone registers as an organ donor, they are not doing it because of a default, or simply because many other people are doing it, but because they believe that it is a good decision in general to help other people and that one does not need the organ when dead. We realise our values by helping someone and also allowing us to help society. It may also work the other way round that we might think it makes sense not to help in this way, because we do not want to reflect on death, or we would like our relatives to make the choice rather than hand it over to unknown professionals. We are not saying we agree with the latter choice, but for agency to make sense, the policy intervention needs us to discover preferences through reasoning, and it must be recognised that a person made a sensible decision for themselves, even if it is not the one the paternalist had originally suggested. Thus, agency-enhancing interventions help in building cognitive capability and motivation through the regular activation of reflective processes.

All our approaches allow us to do this often by targeting cognitive processes that are not automatically in play, so it requires a bit of effort for the individual to get there, at least to begin with. While conventional nudge-type BPP interventions assume those preferences away, (or maybe not, but we just do not know from the delivery of the nudge), these agency-enhancing interventions focus on capacity building and eliciting them explicitly. More active citizens and greater sensitivity, on the part of policy-makers on how citizens are responding to nudges, can create policy signals that then feedback into the process of policy-making and back to citizens again in a virtuous circle, thereby increasing responsiveness of government and citizens. This then leads us to the second advantage of these tools.

These agency-enhancing toolkits foster a mature dialogue between the state and citizens (see Banerjee *et al.*, 2024). We are interested here not in the one-off decisions leading to a behaviour change, but more a continuous set of behaviours underpinned by ideas and attitudes that persist and develop over time, ideally through a dialogue or through ongoing interactions between citizens and the state. Here we can think of BPP, not as a series of nudges, but as part of a virtuous cycle of interventions, reflection, behaviour changes and subsequent feedback from citizens to policy-makers. Nudge has been criticised for relying too much on individual responses, which has been claimed to, in turn, lead BPP astray (Chater and Loewenstein, 2023). Even though BPP, such as over diet and food, had always taken a broad societal perspective, there is still a need for more system-focused interventions to take these social interactions between different parties (such as the citizen and the state) into account (Hallsworth, 2023; Banerjee and Mitra, 2023). Our agency-enhancing devices are better suited to this broader approach to behaviour change and public policy, partly because of the previous advantage, where it improves people's capacities to elicit and articulate true preferences (within bounds), but also because agency takes place in a given context and may change over time. So long as those changes are genuine and not manipulated, and individuals are exercising their choices freely in this more fluid but more productive environment, citizens can engage with governments in ways that will create a more mature public policy dialogue.

Third, agency-enhancing interventions can limit concerns of manipulation that have been raised against conventional toolkits of BPP like nudges (Bovens, 2009; Sugden, 2009; Wilkinson, 2013; Sunstein, 2015; Nys and Engelen, 2017; Sugden, 2018; Schmidt and Engelen, 2020; Ivanković and Engelen, 2022). As discussed earlier, manipulation is always possible when policy-makers hold the reins, which is common for top-down behaviour change interventions like nudges. Building transparency in tools, inherent in our agency-enhancing approaches, allows us to alleviate these concerns (Goodwin, 2012) by engaging citizens in decision-making processes in a more bottom-up approach. For example, when an empowered citizenry realises transparently that the policy-maker is trying to influence them, they can reject such interventions which outrightly undermine their agency. Such empowerment comes only from developing people's agentic capabilities of recognising this manipulation and deciding for themselves, which is an advantage these agency-enhancing interventions offer. Here, it is also important to note that backfiring effects, that is when public policies lead to opposite outcomes than intended, which can arise from such ex-ante transparency in BPP tools, may generate trade-offs between the desired effectiveness and ethics of BPPs. Nonetheless, an active dialogue between state and citizens can minimise these concerns, for example, as shown by Banerjee *et al.*, (2024) who find that individual reflection on a default vaccination enrolment policy in the G7 allowed citizens to realise their intentions and align it with their support for the policy, an effect which is otherwise non-existent in the conventional nudge policy without reflection. Similarly, Diederich *et al.*, (2023) recently showed that while social nudging is more efficient, it is not necessarily ethical. They show that self-nudging helps improve ethics of changing behaviours.

Lastly, these agency-enhancing interventions can lead to a greater uptake of policies, particularly when the goals of the nudgers and citizens match. Here, agency-enhancing interventions can become efficacious for a subset of people who resonate with the directions of the nudge, as decided by the policy-maker or the commercial choice architect. This also has the advantage that BPP can now move towards tailoring interventions for citizens. Recent debate in nudging has shown that there is a wide heterogeneity in the uptake of these policies and that there is little one-size-fits-all. Agency-enhancing interventions harness and build people's own capacities and ensure that citizens can choose policies which they evaluate are best for them. In this way, citizens are given the opportunity to personalise tools of behaviour change. Of course, one can critique this with the obvious argument that paternalism is sometimes in the best interests of citizens (Salanié and Treich, 2009; Conly, 2013). Nonetheless, if that is the case, there is a greater need to question the fit of BPP interventions like nudges in this context. Our tools, discussed above, take a step further here by ensuring a better fit between what governments do and the subgroups that they are talking to. It is partly about eliciting preferences from the interaction, so if the nudge or the policy fails, then another can be denoted which links back to our second argument in fostering a more mature dialogue between these different parties.

There is also the related concern, known as the infantilisation argument (Klick and Mitchell, 2006; Yeung, 2012), that the continual use of nudges can be problematic because they work by harnessing cognitive bias rather than tackling it, so maybe in the longer term erode people's capacities to make conscious decisions and learn good behaviours such as the exercise of self-control on the basis of reflection and judgement (effectively because the nudger is always suggesting the 'correct' way forward). While this is an exaggerated risk we suggest, there is some credence in the position that there are risks associated with placing too much emphasis on approaches which do not build and develop reasoning abilities. In summary, our agency-enhancing frameworks help address many efficacy and ethical challenges identified in some of the critical discussion of nudge (John, 2018; Schmidt and Engelen, 2020; Allcott *et al.*, 2022; Ivanković and Engelen, 2022; List *et al.*, 2022; Maier *et al.*, 2022; Szaszi *et al.*, 2022). We, therefore, call for a greater focus on empowering citizens in designing BPP interventions going forward.

A comparative framework of agency-enhancing BPPs

Having reviewed these alternative frameworks of agency-enhancing BPPs, namely, boosts, debiasing and nudge+, we now turn to a comparative analysis of these toolkits, and identify their shortcomings. We outline four possible limits, which vary across the tools. Table 1 summarises them broadly with a qualitative assessment of how each tool fares against these limitations – with either low, medium, or high risk of this limitation inhibiting the tool's use. These assessments are based mainly on the conceptualisation and design of these toolkits; but where available, we also base these judgements from longer reviews that have been undertaken in assessing these

Table 1. Framework for comparing problems with agency-enhancing BPPs

Limitations	Boost	Debiasing	Nudge+
Cognitive burden	High	Medium	Low
Risk of manipulation – against people's genuine preferences	Low	Medium	High
Psychological reactance – backfire effects	Medium	Medium	Low (if preferences align with nudge goal after reflection), High (otherwise)
Compensatory spillovers	Low	Low	Low

tools. It is important to note that these distinctions are designed to be heuristics designed to start debate about the possible negative impacts of these tools rather than close it off.

The first limitation of agency-enhancing BPPs is the cognitive burden that comes with it. Our basis for these comparisons comes from the conventional nudge that relies on existing biases, so does not take up load because people are processing the information in terms of their own shortcuts, which are automatic and not necessarily conscious. Any attempt to use tools that move away from this must, it would appear, increase cognitive load, because of the time taken to use them. As Table 1 suggests, boosting comes with the highest degree of cognitive load. By definition, boosts in the early stages of behaviour change facilitates learning better competencies, so individuals are required to exert significant cognitive effort in updating their existing 'repertoire of skills'. Nonetheless, once the boost is in play, it reduces cognitive burden as boosted individuals then rely on their smarter new heuristics once again, such as the simple rules of thumb or the good actions bundled with their temptations. On this scale, debiasing requires less cognitive space than boosts as they do not imply undertaking new skill sets, but nonetheless, they rely on purely reflective channels at the time of decision-making. For example, to debias oneself by considering a wider set or options or the mixed frame implies thinking and making comparisons among many elements of a choice set. If people lack information on which to base these judgements, sometimes debiasing may also confuse decision makers, create choice overload and increase levels of difficulty, hence the attribution of 'medium' for this cell. The least amount of cognitive effort is attributed to nudge+ in this set of toolkits. A nudge+ simply asks citizens to evaluate the fit of the nudge with their own goals and preferences. As such, reflection in a nudge+ is narrowly purposed to enable citizens to simply reason their preferences before they accept a nudge. Unlike boosts, they do not require citizens to learn new competencies. They also do not ask individuals to consider a wide range of options as in debiasing. By design, nudge+ interventions put a minimal amount of cognitive load on citizens, partly because the reflective plus tends to be short-lived and easy to undertake.

The next potential limitation of agency-enhancing toolkits is the extent to which they really elicit people's genuine preferences and goals, thereby avoiding risks of manipulation. This after all was at the heart of Thaler and Sunstein's (2003) argument for libertarian paternalism: people often do not and cannot realise what their true preferences are, so that the nudger should take the lead instead. Here, nudges are often accused of making assumptions about people's interests based on limited information about individuals and their differences (Selinger and Whyte, 2011; 2012). Agency-enhancing toolkits, in contrast, appeal to the idea that we are the best deciders of our own interests, a key idea in liberal defences of democracy, reaching back to Stuart Mill and other classic liberal scholars. Yet agency-enhancing approaches also acknowledge that people need assistance in identifying and implementing their true preferences. Here boosts, debiasing and nudge+ each propose different strategies, but have the same goal of how to enable citizens to elicit their own preference. Yet critics may see these agency-enhancing tools as nevertheless susceptible to manipulation by the policy-maker or other choice architect, since they can frame interventions with information that leads to the same end as a nudge. Now the agency-enhancing environment may reassure individuals that they are making an authentic choice, when in fact they are being manipulated, such that the giving of choice makes it easier to get to more repressive or freedom-reducing outcomes (Schmidt, 2017). Marcuse's (1969) idea of 'repressive tolerance' is about how liberal democracy creates a tolerance for a wider range of choice-constraining practices under its umbrella. Seen in this Marxist way, agent-friendly nudges, such as nudge +, may foster this illusion. Since reflection is a means to elicit preferences, reflection tasks can be designed in a way that induce people to submit to the preferences of the nudger. In response to this concern, we would argue that any reflection that precedes or follows a nudge should satisfy conditions that allow individuals to truly develop capacities and report preferences. In this spirit, boosts are least vulnerable to this criticism as they enable people to develop capacities to assess decisions. Of course, much may depend on whether the capacity-enhancing resources are themselves biased to certain world views. Some forms of debiasing may be more vulnerable than boosts to this limitation as the set of options provided to citizens might be chosen with motives that are not in the genuine interest of citizens. In other words, if options chosen are inferior to those being considered by individuals, reflection on them will lead to limited improvements. Similarly, individuals themselves might lack the capacities to make judgements between different options, which then does not overcome the risk of manipulation itself. This concern arises most pressingly with respect to nudge+, which invites the agent to rationalise the prior cause of their behaviour (i.e. the nudge intervention). Such rationalisations, while improving on opaque nudges, are still more open to manipulation than the comparatively more independent type of reflection that boosts and debiasing compel the agent towards. In particular, reflection on the nudge can be framed in ways which are designed to advance the ends of the nudge. So, the nudge continues to unduly influence people's choices, and reflection on it only confirms those choices and, in fact, can even make them stronger.

The third potential concern about agency-enhancing BPP is the possibility of psychological reactance – a perception that choice is being restricted with the consequent resistance to the intervention. Nudges are susceptible to reactance if the subject finds out about the nudge and its objective and does not like being manipulated (Brehm, 1981), so with this cause in mind, agency-enhancing tools should not cause reactance. One way reactance could be caused might be as a consequence of the deeper or more subtle form of manipulation in the above limitation, so that individuals feel more subtly duped as a result, and so reactance might be stronger than for the more obvious nudge. With debiasing, providing information which conflicts with prior beliefs may entrench the original viewpoint, an effect found in some studies of misinformation correction (Lewandowsky *et al.*, 2012). With boosts, people might react against the perceived disparagement of being 'bettered'. Nudge+, contrarily has two distinct effects here. Citizens whose preferences align with the objectives or the directions of the nudge ex-ante are likely to react favourably to the nudge. However, for those where there is a mismatch with these preferences, there is a chance of backfiring if people see the revelation of the aims of the nudge as the state or another actor seeking to limit their autonomy. This was recently shown by Banerjee and Picard (2023) where encouraging people to reflect on social norms promoting sustainable dietary intentions, on average, reduced the uptake of sustainable diets in groups of people who had no ex-ante intentions to change their dietary behaviours. The hope of the nudge+ architect is that people will take the opportunity to revise their ex-ante preferences from reflection, and end at a new more considered ex post preference that better reflects their interests in an autonomous manner, but this can't be guaranteed, nor should it be on agency-promoting grounds. We, therefore, conclude that the psychological reactance to the nudge+ is often an average of these two opposing effects in magnitude. Thus, designing a nudge that aligns with majoritarian preferences will produce less reactance than boosts or debiasing, but higher otherwise.

Finally, behaviour change interventions can cause individuals to respond in ways opposite to the welfare-enhancing directions for a suite of other decisions, as now they feel morally licensed to compensate their good behaviours in one domain with bad behaviours in others (Dolan and Galizzi, 2015; Galizzi and Whitmarsh, 2019). Here, we suggest that none of our agency-enhancing interventions is likely to be negatively affected in this way. This is because, by definition, an agency-enhancing intervention enables citizens to develop their capacities and make choices freely. As such, citizens are not unconsciously primed to undertake good behaviours. Consequently, they do not suffer from warm glow effects arising from the automaticity of their behaviours which lack internalisation of reason and intrinsic motivation. For example, Banerjee *et al.*, (2024) find that neither boosts nor nudge+ interventions to promote sustainable diets reduce participants' contributions to environmental charities afterwards. They conclude there is no evidence of any negative spillovers in this sample of UK respondents. On the contrary, in fact, some of these interventions can lead to promoting (positive) spillover effects. This is common in cases when people realise their true preferences by means of the intervention at play. For instance, if a person reduces their meat consumption by understanding its negative impacts on the environment when they are under the effect of the agency-enhancing intervention, they might choose to undertake more pro-environmental actions afterwards. For example, Banerjee (2022) shows that encouraging UK citizens to reflect on a social norms nudge that promotes the uptake of low-carbon, climate-friendly diets in the UK increases participants' pro-social donation to a charity afterwards, on average. We, therefore, conclude that this risk of inducing compensatory spillovers is low for boosts, debiasing and nudge+.

Conclusion

In the context of BPP, boosts, debiasing and nudge+ are promising interventions but currently underused approaches compared to nudges. These three toolkits described here, as with nudges, have a welfare-enhancing motivation ultimately, seeking to encourage individuals to make choices that their more 'deliberative selves' would select (Thaler and Sunstein, 2008). However the mechanisms to achieve this goal differ, and include training people in the use of helpful heuristics to guide future decision-making (boosting), the use of techniques to reduce instinctive decision-making prone to bias (debiasing), or the encouragement of reflection on attempts to shape people's behaviour (nudge+). All three toolkits, rather than harnessing cognitive biases, counteract it by activating reflective thinking, encouraging reason-based decision-making, and thus ultimately, help put agency into BPP.

The question remains as to how governments can implement such approaches in a practically feasible way while not giving up on the key advantages of nudges, which these approaches can complement. We suggest that, just as governments can mandate for, or against, the use of nudges in the ways suggested by Oliver (2015), they can mandate for the use of debiasing, requiring the use of frames that encourage reflective decision-making, or require 'pluses' to be used alongside nudges in order to make nudges more transparent to citizens who can deliberate over them. These debiasing approaches or nudge+ interventions could become the norm in contexts similar to those where nudges are typically used including food consumption, financial savings, or environmental behaviours. In the case of boosting, investment in education programmes could be targeted at the general population, in primary, secondary or tertiary educational settings, or towards specific groups of professionals such as teachers, clinicians or bureaucrats.

These toolkits recognise the cognitive limitations and systematic biases identified by behavioural economics and sciences without offering either a paternalistic nor a merely libertarian response. Instead, they facilitate consideration of alternatives whilst drawing attention to the issue of concern, leaving the agent to consciously make a decision for themselves. Critically, these approaches help to build cognitive capabilities which may help decision makers learn how to make choices that will best serve their own interest, as defined by the agent themself, both in the immediate moment of the decision but longer term as well, because of the greater transparency, requirement to reflect, and in some cases, the development of decision-making 'skills'. This, in theory, should indirectly benefit society more widely because the individual is being encouraged – albeit through the activation of more reflective processes – to take decisions to maximise longer term welfare but with a transparency not so apparent in BPP approaches to date. We argue that the three toolkits discussed here should be trialled more to help determine whether they can meet the objectives of BPP, which is ultimately about building policy interventions that improve society whilst recognising the important role that psychological processes play in citizen behaviour. Critical questions include the degree of acceptability of these approaches to those at which they are targeted, the risks of backfiring or reactance, their efficacy (for instance, the extent to which they lead to decisions that match the individual's genuine preferences, something which can be retrospectively investigated) and their ability to secure desired public policy outcomes.

References

Allcott, H., D. Cohen, W. Morrison and D. Taubinsky (2022), 'When do" nudges" increase welfare? (No. w30740)', *National Bureau of Economic Research*.

Anscombe, G. E. M. (1957), *Intention*. Oxford: Basil Blackwell.

Arkes, H. R. (1981), 'Impediments to accurate clinical judgment and possible ways to minimize their impact', *Journal of Consulting and Clinical Psychology*, **49**(3): 323.

Audi, R. (1986), 'Acting for reasons', *Philosophical Review*, **95**(4): 511–546.

Banerjee, S. (2022), Choice Architecture 2.0: Can we use reflection in nudges to promote climate citizenship? LSE PhD Thesis.

Banerjee, S. and P. John (2021), 'Nudge plus: Incorporating reflection into behavioural public policy', *Behavioural Public Policy*, 1–16. https://doi.org/10.1017/bpp.2021.6.

Banerjee, S. and P. John (2023a), 'Nudge Plus: Putting Citizens at the Heart of Behavioural Public Policy', in, *Research Handbook on Nudges and Society*, Edward Elgar Publishing.

Banerjee, S. and P. John (2023b), *Nudge+ Think Before you Nudge*. ms, Under consideration by MIT Press.

Banerjee, S. and S. Mitra (2023), 'Behavioural public policies for the social brain', *Behavioural Public Policy*, 1–23. https://doi.org/10.1017/bpp.2023.15.

Banerjee, S. and J. Picard (2023), 'Thinking through norms can make them more effective. experimental evidence on reflective climate policies in the UK', *Journal of Behavioral and Experimental Economics*, 102024.

Banerjee, S., M. M. Galizzi, P. John and S. Mourato (2023), 'Sustainable dietary choices improved by reflection before a nudge in an online experiment', *Nature Sustainability*, **6**(12): 1632–1642.

Banerjee, S., P. John, B. Nyhan, A. Hunter, R. Koenig, B. Lee-Whiting, P. J. Loewen, J. McAndrews and M. Savani (2024), 'Thinking about default enrollment lowers vaccination intentions and public support in G7 countries', *PNAS Nexus*, pgae093. https://doi.org/10.1093/pnasnexus/pgae093.

Battaglio, P. R., P. Belardinelli, N. Belle and P. Cantarelli (2018), 'Behavioral public administration ad fontes: a synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations', *Public Administration Review*, **79**(3): 304–320.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar and S. Galing (2017), 'Should governments invest more in nudging?', *Psychological science*, **28**(8): 1041–1055.

Bovens, L. (2009), 'The ethics of nudge', in , *Preference Change*, Dordrecht: Springer, 207–219.

Bratman, M. E. (1987), *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Brehm, S. S. (1981), 'Psychological reactance and the attractiveness of unobtainable objects: sex differences in children's responses to an elimination of freedom', *Sex Roles*, **7**(9): 937–949.

Brest, P. (2013), 'Quis custodiet ipsos custodes? Debiasing the policy makers themselves', in E. Shafir (eds), *The Behavioral Foundations of Public Policy*, Princeton University Press, 481–493.

Byram, S. J. (1997), 'Cognitive and motivational factors influencing time prediction', *Journal of Experimental Psychology: Applied*, **3**(3): 216–239.

Cantarelli, P., N. Belle and P. Belardinelli (2020), 'Behavioral public HR: experimental evidence on cognitive biases and debiasing interventions', *Review of Public Personnel Administration*, **40**(1): 56–81.

Chater, N. and G. Loewenstein (2023), 'The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray', *Behavioral and Brain Sciences*, https://doi.org/10.1017/S0140525X22002023.

Clarke, R. (2010), 'Skilled activity and the causal theory of action', *Philosophy and Phenomenological Research*, **80**(3): 523–550.

Conly, S. (2013), Against Autonomy: Justifying Coercive Paternalism.

Croskerry, P., G. Singhal and S. Mamede (2012), 'Cognitive debiasing 1: origins of bias and theory of debiasing', *BMJ Quality & Safety*, **22**: ii58–ii64. 2013.

Davidson, D. (1963), *"Actions, Reasons, and Causes", Reprinted in Davidson 1980 Essays on Actions and Events*. Oxford: Clarendon Press. 3–20.

Diederich, J., T. Goeschl and I. Waichman (2023), Self-nudging is more ethical, but less efficient than social nudging.

Dolan, P. and M. M. Galizzi (2015), 'Like ripples on a pond: behavioral spillovers and their implications for research and policy', *Journal of Economic Psychology*, **47**: 1–16.

Drexler, A., G. Fischer and A. Schoar (2014), 'Keeping it simple: financial literacy and rules of thumb', *American Economic Journal: Applied Economics*, **6**: 1–31.

Fischhoff, B. (1982), 'Debiasing', in D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Galizzi, M. M. and L. Whitmarsh (2019), 'How to measure behavioral spillovers: a methodological review and checklist', *Frontiers in Psychology*, **10**: 342.

Godi, M. (2019), 'Beyond nudging: debiasing consumers through mixed framing', *The Yale Law Journal*, **128**: 2035–2086.

Gofen, A., A. Moseley, E. Thomann and R. Kent Weaver (2021), 'Behavioural governance in the policy process: introduction to the special issue', *Journal of European Public Policy*, **28**(5): 633–657.

Goldman, A. (1970), *A Theory of Human Action, Englewood Cliffs*. NJ: Prentice-Hall.

Goodwin, T. (2012), 'Why we should reject 'nudge'', *Politics*, **32**(2): 85–92. https://doi.org/10.1111/j.1467-9256.2012.01430.x.

Grüne-Yanoff, T. (2018), 'Boosts vs. nudges from a welfarist perspective', *Revue d'économie politique*, **128**(2): 209–224.

Grüne-Yanoff, T. and R. Hertwig (2016), 'Nudge versus boost: how coherent are policy and theory?', *Minds and Machines*, **26**: 149–183.

Hallsworth, M. (2023), 'A manifesto for applying behavioural science', *Nature Human Behaviour*, **7**(3): 310–322.

Hallsworth, M., M. Egan, J. Rutter and J. McCrae (2018), *Behavioural Government. Using Behavioral Science to Improve How Governments Make Decisions*. London: The Behavioral Insights Team.

Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and boosting: two distinct pathways to behavior change', *Perspectives On Psychological Science*, **12**(6): 973–986.

Isler, O., O. Yilmaz and B. Dogruyol (2020), 'Activating reflective thinking with decision justification and debiasing training', *Judgment and Decision Making*, **15**(6): 926–938.

Ivanković, V. and B. Engelen (2022), 'Market nudges and autonomy', *Economics & Philosophy*, 1–28.

Jachimowicz, J. M., S. Duncan, E. U. Weber and E. J. Johnson (2019), 'When and why defaults influence decisions: a meta-analysis of default effects', *Behavioural Public Policy*, **3**(2): 159–186.

John, P. (2018), 'The ethics of nudge: Assessing Behavioural Public Policy', in, *How Far to Nudge?*, Edward Elgar Publishing, 108–121.

Klick, J. and G. Mitchell (2006), 'Government regulation of irrationality: moral and cognitive hazards', *Minnesota Law Review*, **90**: 1620–1663.

Larrick, R. L. (2004), 'Debiasing', in D. K. Koehler, and N. Harvey (eds), *Blackwell Handbook of Judgment and Decision Making*, Blackwell Publishing.

Lewandowsky, S., U. K. H. Ecker, C. M. Seifert, N. Schwarz and J. Cook (2012), 'Misinformation and its correction: continued influence and successful debiasing', *Psychological Science in the Public Interest*, **13**(3): 106–131.

Lilienfeld, S. O., R. Ammirati and K. Landfield (2009), 'Giving debiasing away: can psychological research on correcting cognitive errors promote human welfare?', *Perspectives on Psychological Science*, **4**(4): 390–398.

List, J., M. Rodemeier, S. Roy and G. Sun (2022), Judging Nudging: Toward an Understanding of the Welfare Effects of Nudges Versus Taxes (No. 00765). The Field Experiments Website.

Ludolph, R. and P. J. Schulz (2018), 'Debiasing health-related judgments and decision making: a systematic review', *Medical Decision Making*, **38**(1): 3–13.

Maier, M., F. Bartoš, T. D. Stanley, D. R. Shanks, A. J. Harris and E. J. Wagenmakers (2022), 'No evidence for nudging after adjusting for publication bias', *Proceedings of the National Academy of Sciences*, **119**(31): e2200300119.

Marcuse, H. (1969), 'Repressive tolerance', in R. P. Wolff, B. Moore Jr, and H. Marcuse (eds), *A Critique of Pure Tolerance*, London: Jonathon Cape, 95–118.

Mele, A. R. (2003), *Motivation and Agency*. Oxford: Oxford University Press.

Milkman, K. L., J. A. Minson and K. G. Volpp (2013), 'Holding the hunger games hostage at the gym: an evaluation of temptation bundling', *Management Science*, **60**(2): 283–299.

Nagtegaal, R., L. Tummers, M. Noordegraaf and V. Bekkers (2020), 'Designing to debias: measuring and reducing public managers' anchoring bias', *Public Administration Review*, **80**(4): 565–576.

Nys, T. R. and B. Engelen (2017), 'Judging nudging: answering the manipulation objection', *Political Studies*, **65**(1): 199–214.

Oliver, A. (2015), 'Nudging, shoving, and budging: behavioural economic-informed policy', *Public Administration*, **93**(3): 700–714.

Salanié, F. and N. Treich (2009), 'Regulation in happyville', *The Economic Journal*, **119**(537): 665–679.

Schlosser, M. (2019), 'Dual-system theory and the role of consciousness in intentional action', in, *Free Will, Causality, and Neuroscience*, Brill, 35–56.

Schmidt, A. T. (2017), 'The power to nudge', *American Political Science Review*, **111**(2): 404–417.

Schmidt, A. T. and B. Engelen (2020), 'The ethics of nudging: an overview', *Philosophy Compass*, **15**(4): e12658.

Selinger, E. and K. Whyte (2011), 'Is there a right way to nudge? The practice and ethics of choice architecture', *Sociology Compass*, **5**(10): 923–935.

Selinger, E. and K. P. Whyte (2012), 'Nudging cannot solve complex policy problems', *European Journal of Risk Regulation*, **3**(1): 26–31.

Stewart, N. (2009), 'The cost of anchoring on credit-card minimum repayments', *Psychological Science*, **20**(1): 39–41.

Sugden, R. (2009), On nudging: A review of nudge: Improving decisions about health, wealth and happiness by Richard H. Thaler and Cass R. Sunstein.

Sugden, R. (2018), ‘‘Better off, as judged by themselves’: a reply to cass sunstein’, *International Review of Economics*, **65**(1): 9–13.

Sunstein, C. R. (2015), ‘The ethics of nudging’, *Yale J. on Reg*, **32**: 413.

Sunstein, C. R. and R. H. Thaler (2003), *Libertarian Paternalism is Not an Oxymoron*. The University of Chicago Law Review. 1159–1202.

Szaszi, B., A. Higney, A. Charlton, A. Gelman, I. Ziano, B. Aczel and E. Tipton (2022), ‘No reason to expect large and consistent effects of nudge interventions’, *Proceedings of the National Academy of Sciences*, **119**(31): e2200732119.

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

Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Thaler, R. H. and C. R. Sunstein (2021), *Nudge*. Yale University Press.

Wilkinson, T. M. (2013), ‘Nudging and manipulation’, *Political Studies*, **61**(2): 341–355.

Yeung, K. (2012), ‘Nudge as fudge’, *Modern Law Review*, **75**(1): 122–148.

Cite this article: Banerjee S, Grüne-Yanoff T, John P, Moseley A (2024). It’s time we put agency into Behavioural Public Policy. *Behavioural Public Policy* **8**, 789–806. https://doi.org/10.1017/bpp.2024.6