


# Clarifying by declassifying: removing the buzzwords from behavioral public policy

Adam Oliver 

Department of Social Policy, London School of Economics and Political Science, Houghton Street, London, UK  
Corresponding author: A. Oliver, Department of Social Policy, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. E-mail: [a.j.oliver@lse.ac.uk](mailto:a.j.oliver@lse.ac.uk)

## Abstract

Buzzwords abound in behavioral public policy and are used to label various and varied conceptual policy frameworks in the field. The first and most famous of these buzzwords is “nudge,” which in its original manifestation encapsulated a coherent, if limited, perspective. However, instead of acknowledging the limitations of the approach, for which several alternative frameworks were developed to address, the advocates of nudge and those with little expertise in the field widened the parameters of the framework to an extent that its original meaning was largely lost. This essay details these developments and proposes that the buzzwords that are often loosely attached to behavioral public policy interventions—e.g., nudges, nudge-plus, boosts, shoves, and budges—be dropped. Instead, it is suggested that academics, practitioners, policymakers, and the general public reflect more deeply on the type of society in which we collectively wish to live and assess each behavioral public policy intervention on its own terms to discern whether it is congruent with our societal values.

**Keywords:** boosts; budges; nudges; nudge-plus; shoves

At the time of writing, behavioral public policy—i.e., the use of the findings of behavioral science to justify or to use as inputs to inform the design of policy interventions—as a distinct subfield of public policy remains a relatively nascent endeavor, with its origins dating to around 2010. Over its short history, the field has been defined by buzzwords used, in part, to attract attention. For instance, “nudge”—the first and still dominant buzzword in the field—was introduced with an eye toward marketing the book that served as a catalyst for the foundation of the field (Thaler & Sunstein, 2008). Thaler, in his Nobel Prize Lecture, explains that when he and Sunstein “were looking for a publisher for the book we found the reaction to be rather tepid, probably in part because the phrase ‘libertarian paternalism’ does not exactly roll off the tongue. Fortunately, one of the many publishers that declined to bid on the book suggested that the word ‘nudge’ might be an appropriate title. And so, we published *Nudge: Improving Decisions about Health, Wealth and Happiness* [italics added]. In this roundabout way, a new technical term came into social science parlance: a nudge” (Thaler, 2017).<sup>1</sup>

The tactic worked. Although other factors, including a major financial crisis and the accompanying emptying of public sector budgets in many countries that provided fertile ground for those proposing relatively cheap policy solutions, the term “nudge” has entered the common lexicon in the academic,

<sup>1</sup> Libertarian paternalism was the original term that Thaler and Sunstein had used to label their soft paternalistic policy framework (Thaler & Sunstein, 2003). The term was meant to signify their claim that a paternalism that targets automatic decision-making processes does not necessarily erode individual liberty.

policy, and popular discourses. Others have since tried to use the same tactic, albeit with considerably less success, in proposing, for example, the “nudge plus,” “shove,” “boost,” and “budge” approaches (e.g., see Banerjee and John, 2024; Conly, 2013; Hertwig & Grüne-Yanoff, 2017; Oliver, 2013).<sup>2</sup> Unfortunately, in the process of popularizing the term nudge—indeed, perhaps as an unavoidable consequence of popularizing the term—its original, coherent, contestable yet intellectually rigorous definition has been lost. And with that loss, conceptually, the field of behavioral public policy is in something of a mess.

What was the original definition of nudge? For this, I draw my conclusions from Thaler and Sunstein’s 2003 article and 2008 book. The partly interrelated characteristics of the approach are fivefold. (i) That a nudge should improve people’s well-being, as judged by themselves.<sup>3</sup> (ii) That it should preserve liberty, in the sense of allowing people the freedom to choose and behave inconsistently with the outcome objective of the nudge if they wish. (iii) That it does not encompass techniques that are “merely” traditionally rationally informed, such as standard education or information provision and open persuasion. (iv) Similarly, that it does not include interventions that entail significant financial incentives, since such incentives are tools of standard rational choice theory. (v) As a corollary of (iii) and (iv), that it is informed by those robust, systematic aspects of behavioral science—such as present bias, loss aversion and the endowment effect, probability weighting, and anchoring and mental accounting, to name but a few—that are incongruent with the assumptions of the standard rational choice model. As such, information provision and small financial incentives that are explicitly designed with input from these aspects of behavioral science fall within the purview of the nudge approach.

Therefore, in its original manifestation, the nudge approach offered a tight, coherent, if limited, conceptual framework. In short, a nudge was proposed as a nonbinding behavioral-informed paternalistic intervention. Hypothetically, if I expressed a deliberative preference—assuming, of course, that meaningful deliberative preferences can indeed be expressed—to write more articles over the next year, to stop eating after 6 pm every day, and to never check social media again, nudges, in principle, could be designed to try to move me in those directions with the justification that to do so would improve my well-being. What might such nudges look like? They would need to be clearly behavioral-informed and impinge upon my unconscious, automatic decision-making processes, so if a sufficient number of people expressed similar deliberative preferences, the government may, for example, see some legitimacy and promise in issuing posters or leaflets that made salient the losses I (and others) might suffer against some kind of reference point should I (and they) fail in these actions.

However, such interventions might of course have minimal or even no effect (or possibly even a negative effect, if people were to react against them). Indeed, one of the main lines of criticism waged against nudges is that their supposed retention of freedom of choice severely limits their effectiveness. An apparently converse line of criticism is that nudges do not sufficiently respect freedom—or, more precisely, autonomy—over personal lifestyle choices. A third line of criticism, like the first, also suggests that nudges are insufficiently effective at addressing the challenges that contemporary societies face, not because those interventions excessively respect freedom over personal lifestyle choices, but because their targets for behavioral change—namely the demand side rather than the supply side, or in other words citizens rather than, say, industry and corporations—is misplaced.<sup>4</sup> The alternative behavioral public policy conceptual frameworks were developed to address these perceived deficiencies. Unfortunately, the original tight definition of the nudge approach has largely been lost in the professional and popular literatures, and the alternative frameworks have not gained the traction required to bring intellectual clarity to the field. I will argue in this essay that this laxity in definitional standards has resulted in a situation where the terms that have been assigned to these various frameworks serve to obfuscate rather than clarify, and that they should therefore be dropped.

This essay will proceed as follows. In the next section, I will discuss in a little more depth how the original parameters of the nudge approach were gradually widened, such that, for example, simple

<sup>2</sup> These approaches will be discussed in more depth later in this article. Incidentally, although the authors of these alternative approaches have had less success than Thaler and Sunstein in popularizing a term, when one focuses in on the characteristics of each of the policy frameworks there is a strong case to be made that some of these approaches—in particular the boost and budge approaches—have had, and can have, a far more substantive policy impact than nudges.

<sup>3</sup> The “as judged by themselves” condition was introduced in the 2008 book, presumably to counter any notion that the authors were imposing what they and only they believed to be good for others, a common critique of paternalism. Even so, the notion that a person can elicit the genuine deliberative preferences of others has been a source of great contention in the literature (see, for example, Bernheim, 2021; Chater, 2019; Sugden, 2008).

<sup>4</sup> There have been other criticisms; for example, Lodge and Wegrich (2016) suggest that the approach is underpinned by a “rationality paradox,” since there is a disjoint between assuming that those who design and implement nudges rationally act to introduce measures that are premised on the assumption that humans are often irrational.

information provision and nonpaternalistic externality-focused interventions are now often misleadingly labeled as nudges. I will then briefly detail the principal alternatives to the nudge approach, alternatives that were developed to address the limitations of soft, covert forms of paternalism (limitations that nudge advocates—in my view, inappropriately—have come to package under their all-encompassing label). I then present explicitly the various ways in which behavioral science can be used to inform and justify public policy, shorn of any buzzwords, because those buzzwords have often served as a substitute for much needed reflection on whether the parameters of each intervention (i.e., whether they be paternalistic or externality-focused, regulatory or liberty-preserving, or even at their core informed by behavioral science at all) are characteristics that each of us feel are consistent with the essence of a tolerable society. In this essay, to avoid the charge of paternalism (and in acknowledgement of the fact that many people may legitimately disagree with me), I do not wish to impose at length my own vision of what I believe makes for a tolerable society, but in the final section, I feel that it is incumbent on me to offer a few words on my preferred “vision” for the future of behavioral public policy.

## “Nudge” imperialism

Before outlining in a little more depth the main alternatives to nudging, it may be instructive to note how the nudge advocates often dealt with the suggested limitations of their original approach. In my view, they *should* have retained the tight definition of a nudge and defended the approach not as a panacea for most of the challenges that contemporary societies face but as a useful, intellectually coherent approach that paternalists might accept as offering appropriate and effective policy solutions at the margin. Whether or not one is a paternalist and, hence, whether one is ultimately willing to deem nudges as legitimate in practice, one could respect this stated perspective as a reasoned and reasonable proposal. Instead, the advocates of nudging, perhaps driven to retain the status that the nudge label was afforded through first mover advantage and keen to purvey the impression that nudging is synonymous with the whole of behavioral public policy, widened the scope of nudges beyond that which was specified by the original parameters of the approach.

In widening their scope, the proponents of nudge have attached a number of prefixes to their favored term, such that we now have, for instance, “informational nudges,” “educative (or educational) nudges,” “green nudges,” and “social nudges.” Bradt (2022), for instance, in examining hypothetical interventions to increase the demand for flood insurance, designed, among other treatments, a so-called informational nudge.<sup>5</sup> In his study, respondents were asked to imagine that they own a home in coastal USA valued at \$300,000 that has a 1% annual risk of flooding. A flood would cause damages costing \$75,000. The respondents were then asked to consider a flood insurance policy that would cover the cost of these damages. The informational nudge was that the respondents were told the probability of experiencing a flood over a 30-year period (i.e., 26%) and were asked how much they would be willing to pay for the insurance on a sliding scale of \$0–125 per month. Despite the appendage of “nudge,” this intervention is merely the provision of information.<sup>6</sup> In its design, there is no obvious use of behavioral science, if the findings of behavioral science are defined as those robust observations that run counter to the assumptions of standard rational choice theory. Those who might oppose nudges as originally defined due to concerns regarding manipulation are less likely to oppose a simple, explicit provision of information and are left to wonder why the nudge label is used at all in such circumstances.<sup>7</sup>

Informational nudges appear to be a sub-category of educative (or educational) nudges, with the latter also including reminders and disclosures of, say, financial conflicts of interest (see, for example, Reijula & Hertwig, 2022; Sunstein, 2017).<sup>8</sup> But it is not clear that at least some of these types of interventions necessarily find their impact via automatic decision-making processes, as was postulated in

<sup>5</sup> In the examples I cite in this section, my intention is not to single out and criticize any of the authors. In labeling their interventions in the ways that they do, they are simply following what is now common (if unhelpful) practice in the field. I am merely using their examples to illustrate my contention that the use of the nudge label is generally now too lax. I should also note that I do not intend to review the effectiveness of any of the interventions that I mention here. My focus is on their conceptual characteristics.

<sup>6</sup> Calorie counts on menus in restaurants, like road warning signs and maps displayed by car satnav systems, are, for most, familiar pieces of information that are sometimes alluded to as nudges (see, for example, Sunstein, 2019).

<sup>7</sup> If one party deliberately influences the behavior of another party without their knowledge, irrespective of whether or not this is done under the auspices of benefiting the second party (and even if it is claimed that the new behavior is what the second party deliberately desires), then the second party has been manipulated by the first party.

<sup>8</sup> That there is no clear distinction between informational and educative nudges, or indeed often between some categories of those and boosts (to which I will return later), again points to a lack of intellectual clarity in the field.

the original manifestation of the nudge framework. If a financial adviser or institution discloses a conflict of interest, for example, then customers are potentially offered the opportunity to explicitly learn something about this particular service provider that they did not otherwise know, which may in turn impact upon their (more fully informed) decisions.<sup>9</sup> Again, those who are worried about explicitly legitimizing covert public policy instruments in deliberative democracies are likely to be less concerned about mandating many forms of disclosure.

Green nudges target pro-environmental behaviors. While they retain most of the conceptual characteristics of the original nudge approach, they therefore shift the focus of behavior change from that which is adjudged to be best for those targeted to that which is predominantly perceived as best for “others.” In this sense, they are congruent with mitigating negative, or promoting positive, externalities. As noted by Schubert (2017), there are a few reasonably effective interventions of this type: for example, he references work by Allcott (2011; see also Allcott and Rogers, 2014) that shows that leveraging social norms, by regularly reporting to US households how their energy use compares to their neighbors’ use, can reduce energy consumption to some extent. However, Schubert concludes that, generally, there is only limited evidence that so-called green nudges have a substantive effect; moreover, he summarizes various arguments that challenge the ethical acceptability of this type of intervention.<sup>10</sup>

At issue here, though, is neither the effectiveness nor the ethics of green nudges; rather, it is their labeling as nudges (i.e., as forms of paternalism) and the inevitable confusion that this creates. Some may protest that it matters little that they are labeled as nudges so long as they are appropriate and effective, and that they can even be classified as instruments of paternalism if people, on reflection, state that they wish they were doing more to protect the environment.<sup>11</sup> However, mandating, incentivizing, or motivating people to do more for others than they might otherwise do would not normally be thought of as paternalism. Indeed, such externality-driven policy instruments might often be deemed perfectly acceptable by antipaternalists, subject to serious consideration of their relative external and internal costs and benefits. The relaxing of the categorization of what it means to be a nudge to encapsulate very different philosophical outlooks has weakened the intellectual integrity of the approach. Nudge is now “too much,” such that, perhaps paradoxically, the approach, from an intellectual perspective, is almost empty. Many cannot now support or oppose nudge as an approach; if the loose labels are adopted, its acceptance becomes a question of “it depends,” to the extent that to speak of nudge as a general approach to public (and private) policy has become meaningless.

Social nudges are often aimed at the provision of public or collective goods, and therefore their principal focus tends to be on changing behaviors to benefit people other than those targeted for behavior change. Like green nudges, they are not instruments of paternalism and thus fall outside the remit of the original nudge framework. Nagatsu (2015) discusses the ethics of social nudges and maintains that such interventions, when they utilize social norm messaging to influence individual behaviors, do not necessarily undermine autonomy, even if they are directed at unconscious decision-making. Using the simple example of “nudging” to reduce littering, he contends that this is because a social norm message prompts “practical reasoning,” i.e., it provokes the thought that since many others do not litter and I too am expected not to litter, then I had better not litter. However, as already intimated, littering is a harm to others (and, conversely, reduced littering is a benefit to others), and many antipaternalists hold that covert manipulations to discourage such behaviors are, irrespective of any concerns with autonomy, legitimate (although, even here, on grounds of ethical defensibility and effectiveness, they might prefer to endorse more open and explicit intervention). When it comes to the original framework of nudges, one cannot be sure that a person, through their own actions, is really imposing a harm on themselves, and therefore even if we confine ourselves to the use of social norm messaging to address negative externalities, any behavior change that is induced by practical reasoning may in fact be detrimental to the individuals targeted. Moreover, if we place social norm messaging to one side, there are sundry nudge-type interventions where it is more difficult to defend the argument of practical reasoning.

<sup>9</sup> Albeit that disclosure sometimes impacts upon decisions in ways that may seem surprising (see Sah, *Forthcoming*).

<sup>10</sup> In to some extent foretelling the recent hotly debated topic on the relative importance of i-framing versus s-framing in behavioral public policy (see Chater & Loewenstein, 2023), Schubert contends that the individualistic approach of green nudges overlooks the more important socio-cultural roots of environmental harms, which has the potential to detract policymaker attention away from possibly more effective structural interventions.

<sup>11</sup> If the externality is associated with the provision of a public or another collective good, the nudger cannot for sure know that people want to contribute to its financing. The nudger may retort that a nudge does not involve compulsion, but even if one accepts this argument, is it right to allow noncontributors to free ride on the provision of such goods?

To sum up, the main message of this section is that the nudge terminology has been applied to forms of intervention that have little to do with the original, tight, coherent framework that the approach was presented under—that of libertarian paternalism. In short, the parameters of nudge have been widened to the extent that the approach is now almost meaningless. Education and information without explicit behavioral science input and soft forms of externality mitigation are viewed as perfectly legitimate strategies by many antipaternalists, and thus to attach the nudge label to interventions of this type stretches it beyond breaking point and empties it of intellectual content and coherence. It becomes impossible to proclaim whether one is or is not in favor of the approach, unless, of course, one is in favor of almost everything. Luckily, alternative behavioral public policy frameworks, with (less luckily) their own associated buzzwords, have been developed to try to reintroduce some intellectual clarity to the field. These attempts, due in part to the widespread misallocation of the nudge label, have unfortunately met with limited success in this regard, but I will briefly summarize them here.

## Beyond nudging

To reiterate, nudges, as originally defined, were conceptually coherent but limited. They were proposed as instruments of paternalism (i.e., as internality rather than externality-focused), as liberty-preserving,<sup>12</sup> and, irrespective of their advocates' protestations to the contrary, as covert instruments intended to impinge upon automatic decision-making processes. In the years since the introduction of the nudge approach, several other conceptual behavioral public policy frameworks have been developed, in part as responses to what different authors identified as the principal limitations of nudges. These various frameworks have been summarized extensively elsewhere (e.g., see [Oliver, 2015, 2017, 2023](#)) and therefore will not be covered in depth once again here, but brief descriptions of the main ones are warranted.

- (i) Shoves. Although in agreement with the nudge advocates that the behavioral affects can cause people to engage in behaviors that they would deliberately prefer not to do, some critics postulate that by retaining liberty of action, nudges will be insufficiently effective (see [Conly, 2013](#)). Consequently, for behaviors that are judged to be self-evidently bad for people (i.e., actions, such as smoking cigarettes and lack of actions, such as inadequate saving for retirement—actions, so the argument goes, that no reasonable person would condone) should simply be addressed through mandates. For example, smoking ought to be banned, and people should be forced to save more. These instruments fall under a framework of coercive paternalism and are less formally known as shoves.<sup>13</sup>
- (ii) Nudge plus. Others are concerned not so much with the effectiveness of nudges, but with the inexplicitness of their motivation. By relying on their impacting automatic decision-making processes, they argue that nudges potentially undermine people's ability to deliberate. Some authors have therefore proposed that nudge interventions should be accompanied by an explicit explanation of their underlying motivation, so that those targeted for behavioral change can openly deliberate on whether they really want to be nudged. This is the so-called nudge plus approach (see [Banerjee and John, 2024](#)).<sup>14</sup>
- (iii) Boosts. A similar concern with the autonomy-eroding potential of nudges in part led some scholars to postulate that behavioral science should instead be used to improve people's decision-making capacities (see [Hertwig & Grüne-Yanoff, 2017](#)). This can be done by, for example, educating people so as to improve their statistical reasoning abilities, and/or making them aware of the behavioral

<sup>12</sup> Strictly speaking, the advocates of nudging suggest that these instruments preserve freedom—that no options are removed from the table. As intimated at the beginning of this article, while that may be true, nudges inevitably impinge upon the capacity for people to make decisions—i.e., on their autonomy.

<sup>13</sup> Since the advocates of both nudges and shoves postulate that the behavioral affects lead to errors in decision-making that in turn cause individuals to act in ways that they would prefer to avoid, they claim that they are means paternalists rather than ends paternalists (i.e., that people make mistakes with processes rather than preferred outcomes). However, these advocates tend to focus on actions that they themselves appear to approve or disapprove of, and thus, one might contend, particularly with respect to those who postulate shoving, that their attention and concern are directed more toward ends than means.

<sup>14</sup> The nudge plus approach appears to assume that people have the capacity to deliberate on the motivational justification of nudges whenever they encounter one, but for here, the intention is merely to detail briefly the structure of the various behavioral public policy frameworks, rather than the plausibility that they will generally work as intended.

affects that otherwise influence their behaviors unconsciously so that they may use this knowledge to alter their decision-making environments under their own volition as they see fit.<sup>15</sup> These types of intervention are known as boosts, and aside from nudges, are thus far perhaps the most high-profile approach in behavioral public policy.

- (iv) Budes. Although much attention has been devoted to nudges and boosts, it may be contended that an alternative framework, one that focuses on the supply-side rather than the citizen-focused demand-side, is the most ethically defensible and potentially effective of the various behavioral public policy approaches. This approach calls for open regulation against practices that would otherwise be used by self-interested parties to manipulate people into undertaking actions that impose unacceptable harms on them. For example, supermarkets may use the “decoy effect” by placing a product next to a comparable but dominated alternative product (i.e., an obviously “bad deal”) so as to entice people to buy a product that they may not need or even really want. If this is judged to be an unacceptable harm to the consumer, then a regulator has an intellectual justification to act against this type of manipulation.<sup>16</sup> Regulations of this kind, which have been called budes (see [Oliver, 2023](#)), are externality rather than paternalistically motivated, but the externality is not of the traditional kind, where a harm is imposed on a third party to an exchange. Rather, the (external) harm is imposed by one party on the other party in the exchange itself, through a deliberate act that circumvents the awareness of the manipulated party.<sup>17</sup>

As already noted, like nudges, all of these approaches to behavioral public policy have been given buzzwords—i.e., shoves, nudge-plus, boosts, and budes—to define them, and yet a concrete understanding of exactly what these buzzwords define is rare, even among many experts in the field. Despite the best efforts of some over many years to try to introduce some intellectual clarity to the area, that battle is, I fear, lost. Instead of clarity, these buzzwords have been loosely applied and poorly understood (when they have attracted any attention at all), and have for the most part merely served to confuse. Consequently, the time has come to drop their usage from the discourse and instead, when considering policy interventions that are in some sense informed by the findings of behavioral science, to consider each on a case-by-case basis and on their own terms, to reflect on whether their characteristics align with the sort of society in which we wish to live, and with the type of individuals we wish to be.

## Behavioral public policy and the “good” society

More than a century ago, Graham Wallas—one of the four founders of the London School of Economics and Political Science and an early pioneer of behavioral public administration—maintained that when considering policy intervention, one must always keep in mind the question of how to make a “Great Society” (i.e., a large, complex, industrial society, in his day) a “Good Society” (see, for example, [Qualter, 1980](#); [Wallas, 1914](#)).<sup>18</sup> Wallas, although for many years a Fabian socialist, was sympathetic to liberal interventionists, believing that in order to maintain a tolerable level of liberty for all, the liberty of some—namely, those who would otherwise impose unacceptable harms on others—needs to be constrained. Wallas, in believing that people ought to pursue some form of happiness, was something of a moralist, but the question, when considering any form of policy intervention, of “what sort of society do we collectively wish to live in” should always be at the forefront of our minds.

The following is a summary of the ways in which behavioral science can be used to inform policy. For the most part, it is essentially a summary of the frameworks discussed above, but pared of, and thus unhidden behind, any buzzwords.<sup>19</sup> The question we all ought to ask ourselves when considering each

<sup>15</sup> For example, if someone was informed of the power of salience and present bias over individual decision-making, he, or she, might decide to place their store of chocolate at the back, rather than the front, of their kitchen cabinet, if they retain the desire to eat chocolate occasionally but wish to reduce temptation.

<sup>16</sup> This is merely offered as an illustrative example rather than a case where regulation is definitively warranted.

<sup>17</sup> There is a shared concern between this behavioral-informed regulatory approach and [Chater and Loewenstein's \(2023\)](#) argument that the focus of attention in public policy ought to be on the s-frame rather than the i-frame. More generally, the budge approach has been proposed as part of a liberal (antipaternalistic) approach to behavioral public policy, where the focus of attention is on externalities, the mitigation of which can be tackled at both the system and the individual levels (and perhaps other levels besides) (again, see [Oliver, 2023](#)). Chater and Loewenstein are less concerned about the distinction between externalities and internalities. They call for more systems-level intervention, irrespective of whether the intended objective of the intervention is to reduce harms that targeted populations impose upon themselves or upon others.

<sup>18</sup> The other LSE founders were Sidney and Beatrice Webb and George Bernard Shaw. Wallas did not himself use the term a “Good Society,” but he implied that such a society would be one that is tolerable for all (or at least a large majority) of its citizens.

<sup>19</sup> It may be an incomplete list, but it will suffice for the purposes of this essay.



proposed or applied behavioral public policy intervention is how far it corresponds to the characteristics embedded in these frameworks, and how far the characteristics in relation to any specific intervention are tolerable for all, or at least the large majority, of us.

- (a) To use knowledge of behavioral science to manipulate individual behaviors with the intention of specifically benefiting those targeted for behavior change.
- (b) To use knowledge of behavioral science to claim that people are harming themselves and thus to justify enforced behavior change.
- (c) To include an explanation of the reasoning behind any intervention type a, to mitigate the charge that these interventions are manipulating people.
- (d) To use knowledge of behavioral science to make statistical options easier for people to understand and process.
- (e) To inform and educate people about some of the main findings from behavioral science, so that they can use this information to modify their own behaviors if they so wish.
- (f) To regulate against unacceptable behavioral-informed manipulations or for beneficial behavioral-informed manipulations by one party upon another party.
- (g) To use knowledge of behavioral science to manipulate individual behaviors with the intention of benefiting people other than those targeted for behavior change.
- (h) To use knowledge of behavioral science to claim that people are engaging in behaviors that are indirectly harming others to justify enforced behavior change.
- (i) To promote discussion forums and opportunities for collective reflection on specific “irrational” behaviors.

For Wallas, emotions and instincts are powerful and no doubt sometimes distorting influences on individual behaviors that may lead people astray from what they might more deliberately desire to do, but it is difficult—perhaps impossible—for another party (including policymakers) to identify the specific circumstances where mistakes are being made (Wallas, 1908/2010:1914). Moreover, Wallas ultimately did not lose faith in the ability of people to reason. Therefore, of the policy characteristics listed above that are focused upon people’s pursuit of their own desires he may well have favored those in which individual autonomy and agency are most protected (i.e., c, d, e, and i), and favored less those in which people are either manipulated or coerced into altering their behaviors (i.e., a and b). In that he believed that there will be occasions where the liberty of some ought to be constrained to protect the liberty of all, he presumably would not have dismissed approaches that are designed to mitigate the harms that some people impose on others (i.e., f, g, and h).

However, neither Wallas nor any other single individual can rightfully decide what policy characteristics should be favored or avoided to reflect the sort of society in which most of us wish to live. That, on some level, should be a collective decision, which in the liberal democracies must in some sense be an outcome of the democratic process. Within the constraints of this essay, all that can be done is to discuss a few real and hypothetical interventions that can be categorized as examples of behavioral public policy, note which set of policy characteristics they appear to possess, and urge readers to think deeply—to reason—about whether they align with own their view of what makes, or would make, a Good Society.

Let us begin with a policy that was mooted by the UK Government in 2024: namely, a ban on smoking in certain outdoor places, which, if smoking is assumed to be at least in part a consequence of present bias, can be considered a behavioral public policy intervention.<sup>20</sup> If this policy is justified on the grounds that smokers are harming themselves and thus measures ought to be put in place to stop them, then it would align with policy characteristics b. It would be a tool of paternalism and those considering this policy must reason whether any possible health gains (which would be uncertain if people instead decided to smoke at home rather than outdoors) are worth the potential dangers of offering legitimacy to (further) government interference in individual lifestyle behaviors (which may, in turn, snowball into providing further legitimacy over other domains). The question we should ask ourselves (even if we work for the State) is, would an increasing role for the State in interfering in behaviors that have

<sup>20</sup> To reiterate, behavioral public policy covers interventions that are motivated by concerns that behaviors are unreasonably influenced by the behavioral affects, and interventions that are designed specifically with input that comes from knowledge of the behavioral affects (or both). As evidence for the planned smoking ban policy, see: <https://www.bbc.co.uk/news/articles/cg79ym5mrzyo>

few implications for people other than those targeted for behavior change be conducive to the sort of society in which most of us wish to live? If, however, the policy is justified on the grounds of externality mitigation, for instance in relation to reducing passive smoking harms and costs to the National Health Service,<sup>21</sup> then a different set of ethical considerations comes into play. We require plausible evidence that this policy would reduce such harms—e.g., is outdoor smoking really a health hazard to others, and, again, are there likely to be spillover effects that are more harmful than the behaviors that one is legislating against—but assuming such evidence or persuasive arguments are forthcoming, those opposed to this policy on paternalistic grounds may be supportive of it in relation to reducing harms to others. In essence, more might feel that a government that intervenes to limit unacceptable external harms in this regard is consistent with a Good Society. The intervention (i.e., the *same* intervention), due to the altered justification, would now align with policy characteristics h rather than b.

Consider next the tendency for some people to engage in online gambling games to the extent that they place themselves in financial difficulties. If it is assumed that this issue is caused by people often committing to greater risk-seeking as losses mount—which is predicted by certain aspects of behavioral theory (see [Kahneman & Tversky, 1979](#))—then we might conclude that, due to the behavioral affects, they are imposing unreasonable harms upon themselves. Consequently, a view could be taken that particular people ought to be banned from online gambling activities, a proposal that would also align with policy characteristics b. However, policymakers may balk at this intervention if they consider it to be too draconian vis-à-vis a private activity that harms no-one other than the person who gambles, while at the same time expressing concern that many are being unduly manipulated into gambling too much by others (i.e., by the online gambling companies). Therefore, those same policymakers may support, for example, a ban on companies offering “free plays” that entice people to gamble, and/or a regulation that forces the companies to make a reasonable recommended daily spending limit more salient on their platforms (both of these interventions align with policy characteristics f).

If we now turn to something that is perhaps a little more anodyne than gambling debt—littering, for example—a policymaker might, for the good of others, attempt to manipulate rather than regulate would-be litterers by making more salient the positioning of rubbish bins,<sup>22</sup> which would accord with policy characteristics g. Or if this approach was felt to have insufficient teeth, recourse could once again be made to policy characteristics h by increasing the fines for and surveillance of littering. To take another example, policy characteristics b might underpin the enforcement of the wearing of seatbelts, which even some otherwise antipaternalists may support on the grounds of there being little impact of this policy intervention on individual autonomy in return for much potential gain in the reduction of personal injury, and so on.

The policy characteristics thus help us to decide if an intervention is indeed a behavioral public policy intervention, and if so, who exactly it is targeting, for what purpose, and in what way. It is then incumbent on all of us to think deeply about whether each particular intervention is consistent with our own conception of the sort of society in which we wish to live. To do this, we must consider the costs of the intervention in broad terms—not only the direct financial and time costs but also the opportunity costs and the costs associated with reduced autonomy. Moreover, we must consider the possible spillover effects of the intervention, the potential for the ethos underlying the intervention (e.g., State intervention over personal lifestyle choices) to escalate, the regulatory burdens on innovation, and a host of other factors before deciding whether we believe in its legitimacy. It is easy to support an intervention, or even a whole approach, as a “gut” (dare I say it, “behavioral”) reaction, but on reflection, one may often become a little more circumspect.

As noted earlier, no single individual can decide what policy characteristics ought to be favored in any society. But before leaving this essay, I feel that it is incumbent on me to highlight the characteristics that I personally prefer.

## My favored approach

I personally sympathize with two principal messages that arise from the above discussion: (1) Be circumspect of arguments to use behavioral science to justify coercive or manipulative paternalistic

<sup>21</sup> Which are indeed the grounds on which the Government appeared to be justifying the policy.

<sup>22</sup> For example, by painting green footprints on pavements that lead to bins, as observed in Copenhagen: [Pin page \(pinterest.com\)](#)



intervention; (2) Favor the use of behavioral science as an intellectual justification to attempt to mitigate the harms that some people or organizations impose on others.

On point (1), great care ought to be taken when others proclaim that, what is in essence, coercion or manipulation is allowable for the target individual's "own good" (and, in some iterations of paternalism, that this is "as judged by those targeted themselves"). In reality, a third party (i.e., a policymaker) can never really understand what others want from their own lives, because it is very difficult for each of us, with, for example, the myriad of counterfactuals, information asymmetries, opportunity costs, and psychological quirks that we all have and face, to *fully* understand what exactly it is that we truly desire *even for ourselves*. This is not to say that most of us cannot employ reasoning to better understand what we want and to even act upon those deliberations, but we can have no confidence that we know what others want by either inferring our own wants to them or by asking them to try to articulate what it is that they desire. Coercion and manipulation inevitably only tend to align with the desires of the coercer/manipulator rather than those of the coerced/manipulated, and there are clear dangers in legitimizing those approaches. The best proxy for what a person really wants (even though it may often be an imperfect proxy) is observed through the unmanipulated volition of their own actions.<sup>23</sup> This is not to say that there may be instances where most of us, on balance, might support coercive and manipulative paternalistic instruments—for example, as aforementioned, circumstances where the burden of those instruments appears almost negligible and the benefits are (seemingly almost objectively) large. But those instances must never be used to legitimize coercive and manipulative paternalism as *generally* acceptable policy frameworks.

On point (2), I concur with the ethos attributed to Wallas, and predating him, to many others at least as far back as John Locke (1689/2016)—i.e., that there will be circumstances where one needs to constrain the liberties of some in order to protect the liberty of all. This, I believe, is where the focus of behavioral public policy analysts can be most effectively and ethically deployed. In short, it calls for the mitigation of harms to others that are somehow generated by people and organizations who use, implicitly or explicitly, behavioral science phenomena to advance their own egoistic self-interest. Of the policy characteristics summarized in the previous section, it most closely matches those in f (although those in h are also important). This does not necessarily rule out other forms of behavioral public policy, but it does represent what I believe ought to be the main focus of attention if one wishes to engage seriously with the challenges that modern societies face.

A society ruled by those who see it as their duty to consider the harms that each of us may impose upon our fellow citizens rather than the harms that we supposedly impose upon ourselves is the sort of society that aligns with my own system of values. If we reflect on what makes a Good Society, I can only hope that my fellow citizens think deeply not just about what might be gained but what can be lost by broadbrush coercive and manipulative paternalistic intervention by State actors, and ultimately share the perspective that I believe is the essence of liberalism.

## Conflicts of interest

None declared.

## References

- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10), 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>
- Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10), 3003–3037. <https://doi.org/10.1257/aer.104.10.3003>
- Banerjee, S., & John, P. (2024). Nudge plus: Incorporating reflection into behavioral public policy. *Behavioural Public Policy*, 8(1), 69–84. <https://doi.org/10.1017/bpp.2021.6>
- Bernheim, B. D. (2021). In defense of behavioral welfare economics. *Journal of Economic Methodology*, 28(4), 385–400. <https://doi.org/10.1080/1350178X.2021.1988133>

<sup>23</sup> In the field of behavioral public policy, it is assumed that those targeted for behavior change are competent adult agents. Interventions that respect or even improve their agentic competences—for example, education—are of course allowable.

- Bradt, J. (2022). Comparing the effects of behaviorally informed interventions on flood insurance demand: An experimental analysis of 'boosts' and 'nudges'. *Behavioural Public Policy*, 6(3), 485–515. <https://doi.org/10.1017/bpp.2019.31>
- Chater, N. (2019). *The mind is flat: The illusion of mental depth and the improvised mind*. London: Penguin Random House.
- Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 46, e147. <https://doi.org/10.1017/S0140525X22002023>
- Conly, S. (2013). *Against autonomy: Justifying coercive paternalism*. Cambridge: Cambridge University Press.
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986. <https://doi.org/10.1177/1745691617702496>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292. <https://doi.org/10.2307/1914185>
- Locke, J. (1689/2016). *Second treatise of government and a letter concerning toleration*. Oxford: Oxford University Press.
- Lodge, M., & Wegrich, K. (2016). The rationality paradox of nudge: Rational tools of government in a world of bounded rationality. *Law & Policy*, 38(3), 250–267. <https://doi.org/10.1111/lapo.12056>
- Nagatsu, M. (2015). Social nudges: Their mechanisms and justification. *Review of Philosophy and Psychology*, 6(3), 481–493. <https://doi.org/10.1007/s13164-015-0245-4>
- Oliver, A. (2013). From nudging to budging: Using behavioural economics to inform public sector policy. *Journal of Social Policy*, 42(4), 685–700. <https://doi.org/10.1017/S0047279413000299>
- Oliver, A. (2015). Nudging, budging and shoving: Behavioural economic-informed policy. *Public Administration*, 93(3), 700–714. <https://doi.org/10.1111/padm.12165>
- Oliver, A. (2017). *The origins of behavioural public policy*. Cambridge: Cambridge University Press.
- Oliver, A. (2023). *A political economy of behavioural public policy*. Cambridge: Cambridge University Press.
- Qualter, T. H. (1980). *Graham Wallas and the great society*. London: Macmillan.
- Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, 6(1), 119–149. <https://doi.org/10.1017/bpp.2020.5>
- Sah, S. (Forthcoming). The paradox of disclosure: Shifting policies from revealing to resolving conflicts of interest. In *Behavioural Public Policy*. First View.
- Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, 132, 329–342. <https://doi.org/10.1016/j.ecolecon.2016.11.009>
- Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, 19(3), 226–248. <https://doi.org/10.1007/s10602-008-9043-7>
- Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, 1(1), 4–25. <https://doi.org/10.1017/bpp.2016.3>
- Sunstein, C. R. (2019). *On freedom*. Princeton: Princeton University Press.
- Thaler, R. H. (2017). *From cashews to nudges: The evolution of behavioral economics*. Nobel Prize Lecture. <https://www.nobelprize.org/uploads/2018/01/thaler-lecture.pdf>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian Paternalism. *The American Economic Review*, 93(2), 175–179. <https://doi.org/10.1257/000282803321947001>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New Haven: Yale University Press.
- Wallas, G. (1908/2010). *Human nature in politics* (Third ed.). London: FQ Books.
- Wallas, G. (1914). *The great society: A psychological analysis*. London: Macmillan.