*Behavioural Public Policy (2023), 7: 2, 309–332*
*© The Author(s), 2020. Published by Cambridge University Press.* doi:10.1017/bpp.2020.61
First published online 02 December 2020

# Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications

STUART MILLS*
*Department of Psychological and Behavioural Science, London School of Economics, London, UK*

**Abstract:** A recent development within nudge theory is the concept of sludge, which imposes frictions on decision-making. Nascent literature adopts a normative interpretation of sludge: nudge good, sludge bad. However, this normative interpretation leaves much to be desired. A clear definition and treatment of sludge remains absent from this literature, as is a complete understanding of 'frictions'. Furthermore, the relationship between nudges and sludges is unclear. This paper proposes the concept of nudge/sludge symmetry in an attempt to advance the conceptual understanding of sludge. Building from the definition of a nudge, three types of friction permissible under nudge theory are identified: hedonic, social and obscurant. Sludge is then positioned, in terms of frictions, relative to nudge: nudges decrease relative frictions, sludges increase relative frictions. A consequence of this proposition is nudge/sludge symmetry – where a nudge decreases the frictions associated with a specific option, sludge is simultaneously imposed on all other options available to a decision-maker. Nudge/sludge symmetry subsequently challenges the normative interpretation of sludge, and so a new framework drawing on the literature on nudges in the private sector is offered, with the choice architect placed at the centre. This new approach to sludge and emphasis on the role of the choice architect, in turn, reaffirms the importance of transparency in public policy interventions.

Submitted 18 September 2020; revised 28 October 2020; accepted 11 November 2020

## Introduction

Thaler and Sunstein's (2008) concept of nudge has seen remarkable adoption and success in the decade or so since the term was coined. Nudges are often

* Correspondence to: Research Fellow in the Department of Psychological and Behavioural Science, London School of Economics, London, UK. Email: s.mills3@lse.ac.uk



https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

used to help people save for retirement (Madrian & Shea, 2001; Beshears *et al.*, 2016), to encourage healthier food choices (Bucher *et al.*, 2016; Kroese *et al.*, 2016) and to encourage energy-saving behaviour (Allcott, 2011; Allcott & Rogers, 2014), amongst a plethora of other policy applications (Halpern, 2015; Sanders *et al.*, 2018).

A relatively recent development in the world of nudging is *sludge* (Thaler, 2018; Sunstein, 2019, forthcoming; Soman, 2020). Sludge is typically understood as frictions that make good decisions harder (Sunstein, forthcoming), reflecting a normative understanding of sludge that might be summarized as: nudge is good, sludge is bad (Thaler, 2018). Take, for instance, Thaler (2018), who introduces the term 'sludge' into the behavioural science lexicon. Thaler (2018) writes:

> Sunstein and I stressed that the goal of a conscientious choice architect is to help people make better choices 'as judged by themselves'. But what about activities that are essentially nudging for evil? This 'sludge' just mucks things up and makes wise decision-making and prosocial activity more difficult. (Thaler, 2018, p. 431)

This comment, coupled with the concluding remark, 'Less sludge will make the world a better place' (p. 431) and an additional remark by Thaler quoted in Goldhill (2019) – '[Sludge] has two defining characteristics: Frictions and bad intentions' (para. 4) – would certainly suggest Thaler (2018) normatively considers sludge bad.

Thaler (2018) is not alone is this assessment. Ip *et al.* (2018) argue that where nudges should 'nudge us into making better choices without removing our right to choose' (para. 1), 'the goal [of sludge] is different – instead of helping us make better choices, the aim is to unnecessarily increase [costs]' (para. 3). Nobel (2018) also takes this position: '[S]ludge [is] a behavioral intervention that does not have the individual's best interest in mind. It uses the same tools based on cognitive biases and choice architecture, to nudge people towards choices that will not necessarily increase their welfare' (para. 4).

Soman (2020) is not conclusive on the normative question of sludge. Soman (2020) defines sludge as 'frictions in any process that impedes end users, and ultimately reduces welfare' (p. 3), but later writes, 'Sludge impedes our ability to get things done by creating psychological fences. Mind you, not all fences are bad. Sometimes we want to deliberately slow people down from making rash decisions' (p. 3). Thus, Soman (2020) would seem to suggest sludge for good may exist.

Sunstein (forthcoming) offers a broader discussion of sludge. They suggest that where sludge is defined as 'excessive frictions' (p. 4), sludge 'is bad by [this] definition' (p. 5). However, Sunstein (forthcoming) also recognizes that one

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

could choose not to define sludge as 'excessive' frictions, but rather as *increased* frictions (Sunstein, 2019). Thus, 'On [this] definition, sludge would be a kind of nudge – a distinctive subset – and it too could be imposed for good or bad purposes … we could easily imagine "sludge for good"' (p. 7). In doing so, Sunstein (forthcoming), like Soman (2020), entertains a non-normative definition of sludge, just as they do a non-normative definition of *nudge* (Sunstein, 2019).

Current research may benefit from considering sludge from a non-normative perspective, if for no other reason than it is difficult to sustain a normative position in the face of heterogeneity. For instance, even where the net benefits of, say, a nudge are expected to be positive (i.e., welfare-enhancing) on average, and so the nudge may be called 'good', some very heterogeneous individuals may, by virtue of their heterogeneity, suffer as a result of being nudged (Sunstein, 2012; Mills, forthcoming). Equally, what is burdensome sludge for some (or indeed many) may be a valuable shield from the harms of impulsivity for others.

This paper presents a definition and discussion of sludge in line with this second interpretation by considering the (behavioural) frictions that are characteristic of both sludges *and nudges*. Doing so reveals the concept of nudge/sludge symmetry. This concept argues that sludge occurs whenever a nudge is used, and vice versa. Whenever a decision-maker is nudged towards a healthy snack, they simultaneously face sludge if they want an unhealthy snack (Sunstein, 2019). Whenever a decision-maker faces an onerous series of checks to unsubscribe from a magazine subscription (i.e., sludge), they are simultaneously nudged towards keeping the subscription (Soman, 2020). Under nudge/sludge symmetry, both nudging and sludging are defined in terms of relative friction. As a result, the position adopted here is not that sludge itself is a novel development, but that sludge is a novel *reconceptualization* of nudging.

Such a conclusion can only be reached by abandoning a normative position, for if nudges sludge and sludges nudge, one cannot be 'good' while the other is simultaneously 'bad'. This, of course, is messy (though defining 'good' and 'bad' in the first instance is a messy prospect as well). I propose a systematic approach to understanding (good and bad) nudges and sludges by considering how choice architecture is changed and for whose benefit.

The structure of this paper is as follows. Firstly, the concept of nudge/sludge symmetry is developed by considering how nudges work. In understanding the mechanisms that drive nudges, a source of the frictions that are so closely associated with sludge is revealed. So too is the symmetrical relationship between the two. Secondly, the question of normativity is considered. Drawing on previous literature, a simple model of 'good' and 'bad' nudges and sludges (with these terms defined appropriately) is offered. Then, by considering examples of good and bad sludges, the broad normative assumption adopted by some authors and commentators (Ip *et al.*, 2018; Nobel, 2018; Thaler, 2018) is

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

expanded upon. Thirdly, the transparency implications of nudge/sludge symmetry for public policymakers is discussed, followed by concluding remarks.

## Nudge/sludge symmetry

The standard definition of a nudge (Oliver, 2015; Sunstein, forthcoming) is given by Thaler and Sunstein (2008, p. 8):

> [A nudge is] any aspect of choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates.

This is a definition to which I will frequently return. It is immediately interesting to consider how this definition, and nudging generally, can be related to 'friction'. From this definition, nudges cannot significantly change economic incentives. Where economic incentives are significantly changed, such an intervention may more closely resemble a tax, for instance (Oliver, 2015). Yet, this doesn't mean that nudges don't impose *any* significant incentives; nudges, by this definition, can impose non-economic incentives.[1]

For instance, Sunstein (forthcoming, p. 6) writes, '[A] nudge might not impose monetary costs, but it might impose costs of a certain kind.' Sunstein (forthcoming) offers the example of hedonic costs to illustrate this argument, where hedonic costs are broadly taken to be costs that reduce individual pleasure. For instance, Sunstein (forthcoming) argues that a graphic health label may not impose or change the economic incentives of a decision-maker, but the label can certainly induce a degree of discomfort and displeasure for the individual. A similar behavioural cost has been described by Thunström (2019, p. 11) as an 'emotional tax'.

Hedonic costs are not the only explanation of why nudges may work. Within the literature, two additional candidates can be found: (1) social scorn and stigma; and (2) obscurantism.

### *Social scorn and stigma*

Several authors (Bernheim, 1994; Sunstein, 1996; Cialdini & Goldstein, 2004; Cialdini *et al.*, 2005), writing on social norms and conformity, argue that people often adjust their behaviour to fit in with others and avoid social stigma. For instance, Sunstein (1996) argues, 'Many well-known anomalies

1 Hausman and Welch (2010) offer an addendum to Thaler and Sunstein's (2008) definition to incorporate non-economic incentives. I am grateful to Leonhard Lades for this observation.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

in human behavior are best explained by reference to social norms and to the fact that people feel shame when they violate those norms' (p. 909).

Mills (2018) has also contributed to this discussion, noting that the use of significant social incentives in nudging does not violate the concept of liberty as set out by J.S. Mill. For instance, Mill ([1859] 2001) writes, 'For a long time past, the chief mischief of the legal penalties is that they strengthen the social stigma. It is that *stigma which is really effective*' (p. 31, emphasis added). Here, Mill ([1859] 2001) is arguing that a driving force of the law is the social stigma that breaking the law brings. Also consider Tocqueville's ([1831] 2000) discussion of the 'moral power' of majorities (p. 271), or Sætra (2019) for a more contemporary account.

On the use of nudges by choice architects, Mills (2018, p. 401) writes: 'Troublingly, this response [choice architecture] fails to recognise that a constraint need not be weak just because it is non-material. The scorn of my peers may be far more damaging to me than any fine or financial penalty.' Other proposed names for these social incentives are 'moral utility', which Allcott and Kessler (2019) state 'arises when actions impose externalities, are subject to social norms, or are scrutinized by others' (p. 240), and 'social sanctions' (Hausman & Welch, 2010, p. 125).

### *Obscurantism*

The notion of interfering with understanding, either by hiding or by obscuring routes to understanding, has also been discussed as a feature of nudging. Hertwig and Grüne-Yanoff (2017), for instance, argue that the default option nudge may function because it fosters a lack of deliberate understanding. They write: '[S]ome nudges may operate behind the chooser's back and therefore appear manipulative. Default rules can be criticized on these grounds – they take advantage of people's assumed inertia and skirt conscious deliberation' (p. 981). In this instance, Hertwig and Grüne-Yanoff (2017) suggest that default rules seek to utilize people's psychological states, rather than bolster understanding.[2]

A similar notion is offered by Sætra (2020) in their discussion of liberty in the face of big data and nudging. They utilize the term 'psychological force' (p. 2), and they write, 'I dispute the claim that nudging changes behaviour *without the use of force*; the force is simply *psychological*, not *physical*' (p. 6, emphasis in

2 The argument that nudges do not bolster understanding is a common one (Rebonato, 2014; Gigerenzer, 2015). This, largely, is a separate point to the argument being made here. In the language of interference (Sætra, 2020), one can argue that purposely fostering understanding – or *not* fostering understanding – represents a manipulation of understanding regardless of the policy adopted.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

original). Of course, the term 'psychological' is expansive, and insofar as more specificities regarding friction are desired here, a more specific understanding of the term is required. Fortunately, Sætra (2020) offers such specificity, arguing that psychological force describes the purposeful interference in a person's understanding that influences the actions that they take.

Bringing these ideas together, obscurantism may be a good candidate for friction. The Cambridge Dictionary (2020) defines obscuration as 'the act of preventing something from being seen or heard, or something that prevents something else from being seen or heard', while the word itself finds its origins in the Latin for *darkening*. This latter understanding seems more satisfactory than the former, as nudges make concerted efforts not to limit options. But the notions of obscuring understanding (Sætra, 2020) and not promoting understanding (Hertwig & Grüne-Yanoff, 2017) seem resonant with the description of obscurantism as a force for darkening. In some very recent work on sludge, Shahab and Lades (2020) offer the costs of searching for information as a potential candidate for friction; a candidate very reminiscent of what might be called 'obscurant friction'.

## *Frictions*

A tentative taxonomy, incorporating the frictions discussed here, is offered in Table 1. This taxonomy is offered for illustrative purposes and does not make any claim to being exhaustive. These frictions are not far distinguished from the factors that Kahneman (2011) offers as governing an individual's voluntary effort: 'cognitive, emotional, [and] physical' (p. 41). Indeed, these factors map quite well onto the frictions offered here – the exception being social, which one might excuse from a list offered by a psychologist concerned with individual thinking. For instance, cognitive would certainly seem reminiscent of obscurant, physical reminiscent of more classical liberal notions of coercion (i.e., physical imposition and impositions on property) and emotional reminiscent of hedonic costs or emotional taxes.

The proposition offered here is that hedonic costs, social costs and obscurant costs are all prime candidates for what, in discussions of sludge, are called frictions. Take, for instance, paperwork – a popular example of sludge (Sunstein, forthcoming). For an administrator being paid to complete paperwork, while one may debate the efficiency of this expenditure, insofar as the paperwork must be completed, the economic cost of doing so is satisfied.[3]

3 Here, the line between the economic or material costs of sludge and the non-economic and non-material costs of sludge is rather thin. For instance, insofar as an economic value can be placed on time

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

**Table 1.** Frictions.

| Friction | Description | Permissible under nudge theory |
|---|---|---|
| Economic | Changing economic/material (dis)incentives to encourage/discourage specific decision outcomes (i.e., adding a premium charge for a specific outcome) | x |
| Hedonic | Changing individual pleasure/comfort to encourage/discourage specific decision outcomes (i.e., adding a graphic health label to cigarette packaging) | ✓ |
| Social | Changing social/moral costs to encourage/discourage specific decision outcomes (i.e., informing households about the energy use of their neighbours) | ✓ |
| Obscurant | Changing the psychological/cognitive burden to encourage/discourage specific decision outcomes (i.e., using excessive and complicated language in a document) | ✓ |

The concept of paperwork as sludge is the tedium and frustration of it. In this first instance, this may involve processing the same document containing the same information numerous times, invoking hedonic costs, while in the second instance, phrasing, layout or simply an excessive amount of necessary information may make documents frustrating and difficult to understand, invoking obscurant costs. As Sunstein (forthcoming) notes, much of the paperwork undertaken by government could be streamlined by the use of nudges such as default options. This does not eliminate the paperwork per se, but it could make the completion of the paperwork, on average, easier. In this sense, this nudge would reduce the hedonic and obscurant costs borne by those completing the paperwork (say, a claimant of a government provision) and those verifying the paperwork (say, a government administrator). Thus, these costs are the frictions characteristic of sludge. Once more, consider Sunstein (forthcoming): 'Some of the benefits of sludge reduction are psychological and hedonic – a reduction of frustration, anxiety and perhaps a sense

(or insofar as a person's time is a *material* consideration), the somewhat circular logic proposed here that these costs may be justified because someone is willing to pay for them is flawed, hence the acknowledgement of the role of efficiency. But just as this argument is made about sludge, so too is it made about nudge. Sunstein (forthcoming) states: 'If all costs are commensurable, we might be tempted to say that there is no clean line between nudges and material incentives' (p. 6). I defer to Sunstein's (forthcoming) argument that while such arguments may be made, it is useful for the purposes of discussion to *qualitatively* distinguish between various costs (or frictions): 'The only response is that qualitative distinctions are useful, and there is an important qualitative distinction between (say) a tax or a fine on the one hand and a pointless form-filling requirement on the other' (p. 5).

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

of stigma or humiliation' (p. 4). Such frictions have also been identified by Moynihan *et al.* (2015) in their work on administrative burden – a close forerunner of behavioural sludge (Hattke *et al.*, 2019).

But it is important to recognise that the frictions offered in Table 1 are arrived at by considering that which is permissible under *nudging*. If these frictions are exemplar of sludge, they must also be – in some capacity – features of nudging. One explanation is that nudging, generally, should reduce frictions. This is a naturally emerging conclusion: if nudges can reduce sludge as Sunstein (forthcoming) argues, and sludge is understood as an increase in fictions (or indeed, as *excessive* frictions), nudges would seem to reduce frictions.

Armed with this conception of frictions, a new formulation of nudging and sludging is offered:

> Nudges reduce frictions associated with a specified option, while sludges increase frictions associated with a specified option.

### *Symmetry*

Why, then, might the principle of symmetry be offered? The reason is rather straightforward. An understanding of the frictions associated with sludge is arrived at by considering how *nudges* work. If a nudge is reducing the frictions associated with a specified option, the relative frictions associated with all other options are being increased. This is to say, sludge is being imposed as a result of the nudge.[4]

For instance, a default option nudge that automatically enrols employees into a workplace pension scheme reduces the frictions associated with the scheme – a person may face fewer hedonic costs such as anxiety from not saving, fewer obscurant costs as the nudge means they don't have to evaluate all options themselves and fewer social costs as the nudge normalizes saving.[5] But for a person who does not want to be in the scheme, they now face increased frictions, from time wasted opting out, to the cognitive burden of having to understand *how to opt out*, to the social costs of going against the

4 The term 'symmetry' used here may face some objection. After all, don't nudges and sludges impose frictions in opposing directions, and are thus *asymmetrical*? This is a fair comment, though substantially inconsequential. The origin of the notion of symmetry used here is analogous: in particle physics, the concept of symmetry describes particles with the same mass and spin but opposing charges. In the same sense, nudge/sludge symmetry describes behavioural interventions that utilize the same frictions but with differing 'charges', which is to say, either increasing or decreasing frictions.

5 Each potential cost is discussed for completeness. In practice, evidence may suggest some costs are significant and others insignificant.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

grain (Herd *et al.*, 2013). Furthermore, it matters little if this person wants to choose a different pension plan or no pension plan – all other options now face increased frictions, which is to say sludge.

The reverse is also true, hence why this relationship is considered symmetrical rather than asymmetrical. For instance, Sunstein (forthcoming) begins their discussion of sludge with several examples, including immigration procedures, entitlement applications and monetary refunds. In each instance, Sunstein (forthcoming) concludes that, for many people, the onerous nature of these tasks will result in the task, ultimately, being left incomplete. Thaler (2018) suggests similar, characterizing one of the purposes of sludge as being to 'encourage self-defeating behavior' (p. 431).

Thus, the concept of nudge/sludge symmetry is that every nudge imposes sludge on options *not* being nudged towards, while sludge results in relatively easy options being, in effect, nudged towards.[6] Some allusion to this idea has already been made. Sunstein (2019) writes – though only in a footnote – 'If people are nudged to choose healthy over unhealthy food, through good choice architecture, they might face sludge when they seek unhealthy food' (p. 1850, footnote 25). Furthermore, the symmetrical relationship described here fits well with Thaler's (2018) two forms of sludge: 'It [sludge] can discourage behavior that is in a person's best interest … and it can encourage self-defeating behavior' (p. 431). If a decision afflicted with sludge is taken to experience both an encouraging and a discouraging effect on outcomes, I would argue that the *encouragement* is, in terms of frictions, the same as nudging. But to make such an argument, the normative position – nudge good, sludge bad – must be abandoned, as such an argument implies bad, or perhaps 'evil' nudges. I will return to the question of normativity shortly.

## Defining sludge; defining nudge

An important implication of distinguishing between nudges and sludges based on their respective changes to relative (behavioural) frictions occurs when one returns to the definition of a nudge given by Thaler and Sunstein (2008) – this definition says almost nothing of frictions. Furthermore, it would be desirable,

6 The term 'relative' is used as, in many instances, it seems inappropriate to discuss absolute frictions, while offering no clarification would seem to be an oversight. Relativity is proposed from the perspective of a closed decision system (for lack of a better term), where the sum of frictions across the choice set does not change unless the decision system is disrupted (e.g., choices added or removed, new regulations imposed). As such, all frictions associated with one choice are relative to all other frictions associated with all other choices. Of course, there may be a temptation to speak in terms of *absolute* frictions, but this seems practically difficult to do, if not impossible, and certainly contingent on probability estimates.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

having established the notion of nudge/sludge symmetry, to arrive at a definition of *sludge*. A modest alteration to the definition of a nudge given by Thaler and Sunstein (2008) is offered, as well as a complementary definition of sludge:

- A nudge is any aspect of choice architecture that decreases the hedonic, social or obscurant frictions associated with a specific outcome relative to other outcomes and in doing so alters people's behaviour in a predictable way without forbidding any options or significantly changing economic frictions. To count as a nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates.
- Sludge is any aspect of choice architecture that increases the hedonic, social or obscurant frictions associated with a specific outcome relative to other outcomes and in doing so alters people's behaviour in a predictable way without forbidding any options or significantly changing economic frictions. To count as sludge, the intervention must be easy and cheap to avoid. Sludges are not penalties.

One may object to the term 'penalties' used in this definition of sludge, rather than simply mirroring the term 'mandates' used in the definition of nudge. However, following Sunstein (forthcoming), 'A mandate may or may not be sludge, depending on what is mandated' (p. 5). This qualification requires further explanation from Sunstein, as it remains unclear what mandates could be considered sludge within the current expanse of behavioural science. Where less objection can be found is on the question of cost and penalties. Once more, Sunstein (forthcoming) writes: 'If consumers are told that they must pay a specified amount to obtain insurance or that they can obtain a better seat on an airplane for a small additional fee, they are facing costs, not sludge' (p. 5). One perhaps can challenge Sunstein's (forthcoming) use of the adjective 'small' here, as this presents potential difficulties when discussing insignificant economic (dis)incentives, as both definitions imply that small additional fees may be permissible – for instance, a small charge on plastic shopping bags is often taken to be a nudge (or perhaps sludge by the definition given above), but this would also seem not to qualify as a behavioural intervention following Sunstein (forthcoming). Definitional work on what *significant* means remains to be done in nudge theory, but for immediate clarity, penalties are taken here to represent *significant* economic (dis)incentives.

An additional comment on these proposed definitions concerns the frictions named within the definitions. One may prefer the term 'non-economic frictions' rather than the specific frictions given here. These frictions have been included in the definitions as they have previously been disucssed in this article. However, these categories may be disputable – is social scorn not a kind of hedonic cost in terms of pleasure, and where hedonic effects cloud

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

one's judgement, may these not be considered obscurant? In short, these categories are disputable, though an exact taxonomy of frictions is not an intended contribution of this paper. The decision to name specific frictions aims to avoid an unhelpful incidence of tautology – if nudges, *by definition*, do not change economic frictions, to affect some influence they must change some non-economic frictions. In most instances, the specificity of these changed frictions will be of interest, not an exacting list of what these frictions are not. Though, for elegance, I would suggest *behavioural* frictions might suffice when 'non-economic' frictions is undesirable.

## Normative implications

### *An ontological problem*

Nudge/sludge symmetry raises an ontological problem – if every nudge produces sludge, and if every sludge produces nudge, in what language should an intervention be discussed? This is not simply a matter of semantics. As it is hoped that behavioural insights should inform policy (Sanders *et al.*, 2018), the choice of whether to frame an intervention as a reduction in frictions or an increase in frictions could have profound implications on the acceptability of the intervention. For instance, one could imagine, following loss aversion (Kahneman & Tversky, 1979), that a weight loss intervention that is framed as 'making it harder to enjoy guilty pleasures' would be less popular than the framing 'making it easier to improve your health'.

Taking a normative position may appear to solve this problem, but, in reality, it imbues subjectivity. For instance, if a nudge is a reduction of frictions for a 'good' purpose, while sludge is a reduction of frictions for a 'bad' purpose – both subjectively determined – which is the nudge and which is the sludge when opinions on 'good' and 'bad' differ? One might contend that this characterization is inaccurate, as normative sludge doesn't actually *reduce* frictions associated with 'bad' outcomes; instead, it *increases* frictions associated with good outcomes. But this contention does not resolve the issue of subjectivity introduced from adopting a normative position; it merely arrives at a position already adopted here, namely that nudges reduce frictions and sludges increase frictions. In short, because people have differing determinations of 'good' and 'bad', a normative approach offers little recourse to the present issue.

The solution to the problem offered here is a rather simple one: an intervention should be labelled 'nudge' or 'sludge' based on how choice architecture is changed. This position follows from Wendel (2016), who discusses a problem that arises in the digital design community: '[A]ll designs are inherently persuasive' (p. 103). Nudge theorists should recognize this refrain – a common

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

defence of nudging and choice architecture is that choice architecture is unavoidable (Sunstein, 2013). Wendel (2016) argues that interventions should be evaluated based on the following questions: (1) What was developed as part of the intervention? (2) How were these developments applied? (3) What was the intended behavioural outcome?[7]

For instance, automatic enrolment is a nudge because the intervention reduces frictions associated with a specified outcome with the intention of encouraging people to select that outcome. While sludge is imposed on all other choices following nudge/sludge symmetry, this is a by-product of the intervention. By contrast, a subscription service that will only accept cancellation of subscriptions via written notice delivered in the mail is a sludge because the intervention increases frictions associated with a specified outcome with the intention of encouraging an alternative outcome – namely, that people remain subscribed.

### *Pareto and rent-seeking interventions*

If one is to abandon a normative stance on sludge – namely, that sludge is inherently bad – it seems imperative to identify instances of sludge for good.[8] But to do so, one must define quite what is meant by 'good' and, consequently, what is meant by 'bad'. This is no easy feat, given that people are often heterogeneous in their preferences (Sunstein, 2012; Mills, forthcoming). An outcome that is 'good' for one person, as judged by themselves (Thaler & Sunstein, 2008), may be considered 'bad' by another person. Thus, depending on whose perspective is adopted – the former or the latter – any examples come to be interpreted very differently. This is the basis of the weakness of defining sludge normatively.

This is also a weakness of Sunstein (forthcoming), who does much to advance the discussion of sludge for good but does so without establishing criteria for 'good' or 'bad'. As such, I propose a means for determining whether an intervention is good or bad using ideas offered in the literature on nudging in the private sector, where notions of exploitative nudging (what one might call libertarian exploitation) are commonly explored. This framework, of

7 One will note that a similar principle is already embedded within Thaler and Sunstein's (2008) definition of a nudge in that they stipulate that choice architecture should alter behaviour in a *predictable* way, implying both development of an intervention upon which to base a prediction and intentional application in order to test said prediction.

8 I wonder whether the nascent reputation of sludge being bad isn't a product of selection bias. One tends to remember onerous moments of administration or difficulty, even if the ultimate harm was rather negligible. On the other hand, does one tend to recall moments of ease that lead to tremendous harm, or instead focus on the harm without a second thought for the mechanism by which one stumbled into it?

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

course, is not beyond criticism, as any attempt to define in *any* solid form the meaning of words such as 'good' and 'bad' only invites Wittgenstein's monster. I would simply argue that any specification for determining how 'good' and 'bad' should be understood is better than none.

Bar-Gill (2012) argues that many commercial contracts are designed so as to nudge consumers towards outcomes that may not be of benefit to them, but the choice architect (i.e., the company). Because of this, Bar-Gill (2012) argues that government has a role in providing counter (welfare-promoting) nudges to protect consumers.

In another discussion of nudging in the private sector, Beggs (2016) proposes 'Pareto' and 'rent-seeking' nudges (p. 127). Pareto nudges are to be understood as nudges where both the decision-maker (i.e., the consumer) and the choice architect (i.e., the company) benefit from the outcome being nudged towards. The term 'Pareto' is borrowed from the economic lexicon but is not used to describe a Pareto optimality. Pareto here considers only the choice architect and the decision-maker, and not third parties, who of course may or may not benefit directly or indirectly as a result of an intervention. Rent-seeking nudges, by contrast, are to be understood as nudges where only the choice architect benefits from the outcome of the nudge.[9]

Note that these conceptions of nudging in the private sector are not in opposition. The nudges that Bar-Gill (2012) considers to require counter-nudging by government would seem to be rent-seeking nudges, while one can interpret these counter-nudges by governments as being Pareto insofar as choice architects within government also benefit from the nudge. This is not an unreasonable assumption; good policy benefits politicians seeking re-election and benefits public servants seeking promotion (Rebonato, 2014).[10] Furthermore, choice architects are often biased themselves (Rebonato, 2014), and as citizens who themselves can be nudged, they can also reap the benefits of Pareto nudges. Finally, this framework of Pareto and rent-seeking

9 Benefit is taken here to broadly mean welfare, which follows the general treatment of nudges in the literature (Allcott & Kessler, 2019; Sunstein, 2020) and in how Beggs (2016) deploys these terms.

10 This is something of a reversal of Rebonato's (2014) argument. Rebonato (2014) criticizes nudging by government because '"government" is not an abstract entity (benevolent or malevolent, according to one's political bent). It is made up of individuals, with their own interests (first and foremost to be elected and reelected), who operate through a bureaucracy which is in turn made up of real people with their agendas, interests, biases and, yes, bounded-rationality limitations' (p. 360). However, within a well-functioning democracy, the motivations of these individuals who assume the role of choice architect should align with the interests of (most) voters (Downs, 1957). Of course, they will never align with all individuals because individuals are heterogeneous (Sunstein, 2012), but the same motivations that Rebonato (2014) characterizes as the flaws of choice architects can also be argued as motivating conscientious choice architecture.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

nudges captures quite accurately the notion of 'bad' intentions discussed in the nascent sludge literature.

Thaler (2018) argues that sludge is bad or 'nudging for evil' (p. 431) because it makes prosocial behaviour more difficult. To be sure, Pareto nudges are different from prosocial behaviour insofar as Pareto nudges focus on reciprocity while prosocial behaviour is concerned with the benefits of others.[11] But where the benefits for others produce a benefit for oneself – which, it is argued here, is the case for both the public and private sector – this discrepancy is insignificant. Ip *et al.* (2018) – who focus on sludge in the private sector – argue that the purpose of sludge is to raise costs for decision-makers. This is not inconsistent – and in fact seems rather aligned with – Beggs' (2016) notion of rent-seeking nudges and Bar-Gill's (2012) critique of private-sector nudging. Finally, Nobel (2018) argues that sludge is a nudge that does not have the decision-maker's best interests in mind. As discussed above, this equivalency between sludge and nudge is rejected, but the implicit understanding of why sludge is bad – it does not have the best interests of the decision-maker in mind (Soman, 2020) – can easily be captured within the concept of a rent-seeking intervention. Even above, in the discussion of sludge as paperwork, the benefits of reducing sludge through nudging were described in terms of a Pareto nudge – applicants face less of a burden in completing paperwork and administrators face less tedium and challenge in verifying paperwork.

The terms 'Pareto' and 'rent-seeking', therefore, may be a useful basis from which to construct a proposition given the present normative difficulties. Adapting Beggs' (2016) terms slightly, a Pareto intervention is taken to be an intervention where the choice architect may maximize their own benefit by maximizing the benefit received from the intervention by the decision-maker. In some instances, the maximal benefit to one party may be no change in benefit. A rent-seeking intervention, by contrast, is an intervention where the maximum benefit for the choice architect can be achieved by using an intervention that does not maximize the benefit for the decision-maker. In

11 This assertion may be a point of contention either as some may not define prosociality as benefit *exclusive* to others or because some may consider reciprocity prosocial. On the latter, for instance, Oliver (2019) argues that reciprocity – which the Pareto nudge idea integrates – is 'substantively prosocial' (p. 922), and thus one may choose to replace the terms 'Pareto' and 'rent-seeking' with 'prosocial' and 'pro-self', respectively. The substantive benefit of this, however, seems only to reduce the use of esoteric language. Furthermore, this is possibly a curse in disguise – conceptions of prosocial and pro-self nudges already exist within the literature (Hagman *et al.*, 2015; Tyers, 2018), and thus using these terms here may simply confuse matters. Another issue regarding reciprocity may be the implication of some interaction between the decision-maker and the choice architect, which may be an unfair assumption in some contexts. The simpler idea of mutual benefit may, therefore, be more appropriate. I am grateful to a reviewer for this suggestion.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

some instances, both parties may benefit from a rent-seeking nudge, but the decision-maker's benefit would not be maximized, even when increased.

The common conceptions of bad sludge (which often, in terms of frictions, describes bad *nudges*) can be understood consistently using the language of Pareto and rent-seeking interventions. This language offers another advantage – an intervention is determined to be Pareto or rent-seeking based on the *incentives motivating the choice architect*. Thus, there is a degree of parsimony in this conceptual approach: when delineating whether an intervention should be called a nudge or a sludge (assuming it is one of these two), it is argued that the change to friction made by the choice architect should be evaluated; when delineating whether an intervention is 'good' or 'bad', it is argued that the incentives for the choice architect as they relate to decision-makers should be evaluated. The choice architect is central in both instances.

Henceforth, Pareto and rent-seeking interventions are taken to define 'good' and 'bad' interventions, respectively. An intervention is described as good if both the choice architect and the decision-maker would be expected to benefit. An intervention is described as bad if the choice architect is expected to benefit while the decision-maker is not.

## Sludge for good

Having established on what basis 'good' and 'bad' will be determined (i.e., Pareto and rent-seeking, respectively), attention may now turn to the question of sludge for good. Sunstein (forthcoming) offers two examples, namely (1) cooling-off periods, and (2) are-you-sure checks. Furthermore, Sunstein and Gosset (forthcoming) introduce another potential candidate for good sludge – administrative burdens that reduce the false or fraudulent claiming of government benefits. The latter, however, as Sunstein and Gosset (forthcoming) make clear, is often a balancing act, and can easily become damaging when frictions impede too many qualifying applicants from accessing their benefits. For this reason, the proposition by Sunstein and Gosset (forthcoming) is not discussed further. Also see Thaler's (2018) discussion of sludge and voter fraud.

In addition to the examples of cooling-off periods and are-you-sure checks, I propose a third example: disfluency.

### *Cooling-off periods*

Cooling-off periods refer to a period of time (typically) following a purchase in which a person may reconsider their decision without any consequences beyond those stipulated under the conditions of the cooling-off period. Cooling-off periods may also apply in areas such as higher education, where

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

students may be able to change their major/course of choice within a given period of time, or divorce law (Soman, 2020; Sunstein, forthcoming).

For vendors who must implement cooling-off periods, the benefit seems limited, as the period gives customers time to request a refund that the vendor would rather not be obliged to provide. However, in many instances, it is not the vendor but the government who mandates cooling-off periods in various circumstances, such as in a mortgage application.

From the perspective of government as the choice architect, a Pareto relationship can be identified, with the government seeking to foster favour with the public by providing decision-makers with decisional protection in the form of a cooling-off period. In the instance of, say, a university allowing students time to evaluate their subject choices – even when this is not mandated by the state – an expectation of mutual benefit by the choice architect is reasonable, as a student may appreciate the flexibility that a cooling-off period provides when making such a large life decision, and the university may benefit from a student who is still satisfied and in attendance. Insofar as companies seek to maintain a good reputation with their customers, companies may also find a cooling-off period to be of mutual benefit. These examples, then, may also be described as Pareto relationships.

However, one may argue that cooling-off periods are a rather dubious form of sludge, as most periods do not end with a vendor asking the decision-maker if they wish to change their decision – decision-makers are merely assumed to be satisfied. In this sense, a cooling-off period is sludge only insofar as the frictions it imposes represent a prompt to re-evaluate a decision and an opportunity to choose again.

### *Are-you-sure checks*

What may render a cooling-off period more typically sludge is the addition of an are-you-sure check (Soman *et al.*, 2010). Such checks prompt decision-makers to evaluate if they are *really* sure that they want to choose a given option before ultimately committing.

The parallels between are-you-sure checks and cooling-off periods are rather obvious, therefore, though the are-you-sure check takes the sludge a bit further. Unlike the cooling-off-period, which has an implicit default that a customer is satisfied with their decision, the are-you-sure check has no default and demands the customer make a confirmatory choice (Soman *et al.*, 2010). In this sense, the are-you-sure check imposes more frictions than the cooling-off period, and where such a check is mandated, a candidate for the 'mandates as sludge' proposed by Sunstein (forthcoming) may be revealed.

The use of are-you-sure checks has become popular on social media in recent years, particularly following the rise of popular misinformation and the risk of

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

personal-life consequences of social media posting. On the former, the social media site Twitter has recently implemented a feature asking users if they are sure that they want to share a link to content they themselves have not read (Hern, 2020; Soman, 2020), while on the latter, the photo-sharing site Instagram now prompts users if they are sure that they want to share a comment that is potentially hurtful (Bryant, 2019). These prompts are sludge and have even been described within the language of sludge and nudge. Hern (2020), for instance, writes, 'Twitter's solution [to sharing unread information] is not to ban such retweets, but to inject "friction" into the process, in order to try to nudge some users into rethinking their actions on the social network' (para. 6).

In the case of social media, the frictions imposed appear to be social (e.g., 'are you sure you want others to see this?'), but one could also imagine the imposition of hedonic costs in other contexts, such as weight loss (e.g., 'are you sure you want dessert while you're dieting?'). As with cooling-off periods, the mutual benefit of this sludge can be seen. For instance, social media sites prosper on the quality of content supplied by users of their platforms. If content is seen as being substandard or hurtful, the platform will suffer. Equally, users who consume said content benefit not only from avoiding the potential consequences of their own posting, but also from a better standard of content and discourse generally.

### *Disfluency*

Perhaps the most interesting instance of sludge for good can be found in the use of purposely difficult, yet still conscientious, fonts and aesthetics. This follows the notion of *disfluency* – the act of making tasks more cognitively difficult.

Diemand-Yauman *et al.* (2011) demonstrate the power of disfluency in *increasing* understanding by using text fonts that are harder to read (i.e., bad fonts). The authors report that the added difficulty of reading meant that readers had to engage with the text more, and thus this promoted their understanding of the content. As Benartzi (2017) writes, 'Sometimes, people actually remember more when the information is slightly harder to process; the perceptual struggle is a good thing' (pp. 122–123). Benartzi (2017) even contrasts disfluency with nudge, writing, 'Richard Thaler puts it [nudge] this way … "Make it easy." … But making things easier is not always ideal' (pp. 121–122). It seems likely that Benartzi (2017) would have related disfluency to sludge had the concept been established at the time of their writing.

For instance, disfluency imposes obscurant frictions that, in the vein of Sætra's (2020) psychological force, interfere with understanding. But unlike the common perception of difficulty *reducing* understanding, Diemand-

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

Yauman *et al.* (2011) report the opposite. Insofar as understanding may be crucial, such as during a mortgage application or car rental (Benartzi, 2017), the use of disfluent techniques such as bad fonts can be considered sludge for good. Kahneman (2011) likewise reports on a type of disfluency when using a second language; the lack of immediate intuition slows thinking down and can lead to more considered outcomes. Even in marketing, evidence of the mutual benefit of disfluent fonts has been identified. Motyka *et al.* (2016) find significant evidence that disfluent fonts increase customer engagement with promotional offers, leading customers to choose the better option. Thus, where a vendor is seeking to promote a product that is better value for the customer than an alternative, more fluent brand, the added sludge of a difficult font may lead to mutual benefit.

Furthermore, *fluency* can be deceptive and induce bias. As Carpenter *et al.* (2013) find, fluent instructors often left students feeling as though they had learned a lot, while actual assessments of learning showed no significant benefit. To reiterate Benartzi (2017), 'making things easier is not always ideal'.

## *The quirks (and quarks) of nudge theory*

There is an inconsistency across Sunstein (2019) and Sunstein (forthcoming) that complicates the discussion presented in this paper. Sunstein (2019) writes, 'It should be clear that nudges can be for good or for bad. It should also be clear that sludge can be for good or for bad' (p. 1850, footnote 25). From this statement, the normative position set out above – nudge good, sludge bad – is not adopted by Sunstein (2019). On the other hand, Sunstein (forthcoming) writes, '[W]e can see that some helpful nudges reduce frictions ("make it easy"), while other helpful nudges increase frictions ("make it hard")' (p. 7). Here, Sunstein (forthcoming) labels two functionally different interventions (in terms of frictions) as nudges, seemingly on the basis that both functions prove 'helpful'. The normative position has returned.

This is not to accuse Sunstein (forthcoming) of error; when defining sludges merely as 'frictions', Sunstein (forthcoming) argues that 'sludge would be a *kind of nudge*' (p. 7, emphasis added), thus resolving this surface-level inconsistency. Yet, this all still remains messy, which is why I propose that the systematic approach of understanding (good and bad) nudges and sludges from the actions and motivations of the choice architect is desirable.[12] This approach does not require the various caveating of nudges as seen with, as

12 I am reassured by Sunstein's (2019) proclamation that, 'to be sure, more work remains to be done on definitional issues', followed by their useful comment, 'My hope is that the examples [provided here] will be sufficient for purposes of the current discussion' (p. 1850, footnote 25).

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

**Table 2.** Good and bad, nudge and sludge.

| | 'Good' | 'Bad' |
|---|---|---|
| Decreased frictions | Pareto nudge | Rent-seeking nudge |
| Increased frictions | Pareto sludge | Rent-seeking sludge |

Sunstein (forthcoming) dubs them, *deliberation-promoting nudges* – these may simply be dubbed Pareto sludges (see Table 2).

However, Sunstein (forthcoming) explains the reasoning for the caveat – namely that sludge is generally understood as an unpleasant entity and thus it may be beneficial to reserve such a term only for unpleasant instances and refer to 'sludge for good' in terms of nudging. This, I would argue, is not a robust argument, for two reasons. Firstly, pejorative connotations should not supersede substance – I use the terms 'nudge' and 'sludge' based on the change in relative frictions, not as a reflection of whether an intervention is good or bad given subjective determination. Secondly, could not the same pejorative argument be made about the term 'nudge'? On this basis, this whole discussion devolves into an unhelpful debate over language.

## Implications for public policy

Nudge/sludge symmetry, and indeed sludge as a novel reconceptualization of nudge, may create challenges for public policymakers. One such challenge has already been discussed – what might be called the 'branding' challenge of behavioural interventions. Where nudges sludge, and sludges nudge, the choice of how to frame an intervention to the public is one of consequence.

This challenge is partially informed by another challenge that is not a novel imperative identified here, but is certainly given greater importance following the arguments made above – transparency.

Transparency, or the lack thereof, has long been a criticism of nudges (Bovens, 2009; Lades & Delaney, forthcoming). For instance, Rebonato (2014) argues that nudges are intentionally opaque and therefore struggle to justify any claims to preserving genuine freedom of choice. While this claim has partially been refuted by an expanding body of literature (Loewenstein *et al.*, 2015; Steffel *et al.*, 2016; Bruns *et al.*, 2018; Bang *et al.*, 2020),[13] the

13 This body of research focuses primarily on what would be described as Pareto nudges. It may be a valid hypothesis that the effectiveness of rent-seeking nudges in the presence of transparency would not fare as well.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

importance of transparency within nudge theory has remained, insofar as one questions the *reasoning behind the use of behavioural interventions* (Lades & Delaney, forthcoming).

It is important that it should be possible to scrutinize nudges (Delaney, 2018). Thaler and Sunstein (2008) argue that such scrutiny should follow Rawls' (1971) publicity principle, whereby a policy must be sufficiently transparent so as to be understood by those it will affect and rejected if necessary. An important principle espoused in this paper – that the choice architect is central – only reinforces this imperative. In the first instance, such that the framing of an intervention (i.e., nudge or sludge?) can be properly critiqued, choice architects should be transparent as to what choice architecture they have changed. In the second instance, choice architects should be open about their motivations and the motivating factors behind any behavioural intervention, providing – where possible – explanations of the expected benefits of an intervention and – when dealing with heterogeneous groups – explanations for their selection of said groups (Mills, forthcoming). Recent literature (Reijula & Hertwig, forthcoming) has even begun to suggest that transparency regarding nudging should extend beyond merely revealing the presence of an intervention to revealing the psychological mechanisms and biases that enable nudges to work.

Of course, beyond merely discussing what elements of choice architecture should be made transparent, it is a pertinent question to whom transparency information should be made available. The publicity principle provides a valuable guide in this regard – in a world of intelligently designed nudges and sludges, transparency regarding (1) the motivations, (2) the calculations, and (3) the mechanisms of the interventions should be available to all of those impacted by the interventions, which it is argued here is most likely all citizens.

Furthermore, nudge/sludge symmetry lends credence to an idea previously proposed by Sunstein (forthcoming): that of the sludge audit. If, as it is argued here, all nudges produce sludge and all sludges produce nudge, it seems reasonable to suspect that a tremendous amount of decision interactions contain a wide variety of frictions that, through intelligent choice architecture, could be improved upon. Sunstein's (forthcoming) sludge audits are designed to 'catalogue the costs of sludge and to decide when and how to reduce it' (p. 1). However, given nudge/sludge symmetry, instances of rent-seeking nudges seem as worthy of resolution by a conscientious choice architect as that of rent-seeking sludges, and so a broader notion of a *behavioural audit* might be emphasized.

Returning to the question of transparency, behavioural audits could function as accessible reports produced by choice architects, both public and private, and would likely include, beyond cost–benefit evaluations as proposed by

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

Sunstein (forthcoming), declarations of interest on behalf of choice architects, internal evaluations of the efficacy of various behavioural interventions and a summary of the audience that choice architects *think* their interventions are targeting.

Finally, any transparency in behavioural interventions – whether via a behavioural audit or another framework (Lades & Delaney, forthcoming) – should consider the relationship between obfuscation and transparency. The goal for transparency, particularly in the matters of interpretation that nudge/sludge symmetry invokes, should be such that disclosures can be interpreted by others (e.g., decision-makers, other policymakers, citizens generally) in a meaningful way, rather than a mass of disclosure information being used to obfuscate understanding (Berg, 2018; Mersch, 2018).

## Conclusion

This paper offers a contribution to the nascent sludge literature by proposing the notion of nudge/sludge symmetry. Through a consideration of the drivers of nudging, three candidates for 'friction' often associated with behavioural sludge are identified: hedonic frictions, social frictions and obscurant frictions. With this understanding of friction, I reconsider the relationship between nudge and sludge, arguing that the former decreases frictions associated with a specific option, while the latter increases frictions associated with a specific option. Given such a relationship, it is argued that nudging, as a by-product, increases relative frictions on all other options (i.e., sludge), while sludging, as a by-product, decreases relative frictions on all other options (i.e., nudge). This is nudge/sludge symmetry.

Nudge/sludge symmetry challenges the normative position of 'nudge good, sludge bad'. Under nudge/sludge symmetry, any 'good' nudge also imposes 'good' sludge, which is not an acceptable conclusion under this normative position. As such, I reconsider the role of normativity in nudge (and sludge) theory and argue that, by defining nudges and sludges in terms of (changes in) frictions, the normative position must be abandoned.

Insofar as it is useful to talk of 'good' and 'bad' interventions, however, abandoning a normative (and implicitly subjective) position creates issues. As a resolution, I draw on the private-sector nudging literature and utilize Beggs' (2016) notions of Pareto and rent-seeking nudges to establish a criterion for the appraisal of the goodness or badness of an intervention. Nudge/sludge symmetry and this new criterion re-emphasize the centrality of the choice architect and, I argue, in turn re-emphasizes the importance of transparency in public policy nudging/sludging.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

## Acknowledgements

I am grateful to Richard Whittle, whose project led me to write this paper. I am also grateful to Kevin Albertson and Leonhard Lades for their kind and helpful comments on an early draft, and to Henrik Skaug Sætra for sharing preprint materials with me. Finally, I am grateful to the two anonymous reviewers for their kind and helpful comments. All errors are my own.

## References

Allcott, H. (2011), 'Social norms and energy conservation' *Journal of Public Economics*, **95**: 1082–1095.

Allcott, H. and J. Kessler (2019), 'The Welfare Effects of Nudges: A Case Study of Energy use Social Comparisons' *American Economic Journal: Applied Economics*, **11**(1): 236–276.

Allcott, H. and T. Rogers (2014), 'The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation' *American Economic Review*, **104**(10): 3003–3037.

Bang, H. M., S. Shu and E. Weber (2020), 'The role of perceived effectiveness on the acceptability of choice architecture' *Behavioural Public Policy*, **4**(1): 50–70.

Bar-Gill, O. (2012), *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*, UK: Oxford University Press.

Benartzi, S. (2017), *The Smarter Screen: Surprising Ways to Influence and Improve Online Behavior*, USA: Portfolio Penguin.

Berg, J. (2018), 'Obfuscating with transparency' *Science*, **360**(6385): 133

Bernheim, D. (1994), 'A Theory of Conformity' *Journal of Political Economy*, **102**(5): 841–877.

Beshears, J., K. Milkman, H. Dai and S. Benartzi (2016),, '*Framing the Future: The Risks of Pre-Commitment Nudges and Potential of Fresh-start Messaging*' Working Paper. [Online] [Date Accessed: 09/04/2019]: https://static1.squarespace.com/static/5353b838e4b0e68461b517cf/t/583ca5acd2b8571174b28e40/1480369581625/48-Beshears_et_al_2016.pdf.

Bovens, L. (2009), 'The ethics of nudge' in T. Grüne-Yanoff and S. O. Hansson *Preference Change: Approaches from Philosophy, Economics and Psychology*, (2009), Berlin: Springer.

Bruns, H., E. Kantorowicz-Reznichenko, K. Klement, M. L. Jonsson and B. Rahali (2018), 'Can nudges be transparent and yet effective?' *Journal of Economic Psychology*, **65**: 41–59.

Bryant, M. (2019), '*Instagram's anti-bullying AI asks users: "Are you sure you want to post this?"*' The Guardian. [Online] [Date accessed: 03/08/2020]: https://www.theguardian.com/technology/2019/jul/09/instagram-bullying-new-feature-do-you-want-to-post-this

Bucher, T., C. Collins, M. Rollo, T. McCaffrey, N. de Vlieger, D. van der Bend, H. Truby and F. Perez-Cueto (2016), 'Nudging consumers towards healthier choices: a systematic review of positional influences on food choice' *British Journal of Nutrition*, **115**: 2252–2263.

Cambridge Dictionary (2020), 'Obscuration' [Online] [Date accessed: 18/09/2020]: https://dictionary.cambridge.org/dictionary/english/obscuration

Carpenter, S., M. Wilford, N. Kornell and K. Mullaney (2013), 'Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning' *Psychonomic Bulletin and Review*, **20**: 1350–1356.

Cialdini, R. and N. Goldstein (2004), 'Social Influence: Compliance and Conformity' *Annual Review of Psychology*, **55**: 591–621.

Cialdini, R., L. Demaine, B. Sagarin, D. Barrett, K. Rhoads and P. Winter (2005), 'Managing social norms for persuasive impact' *Social Influence*, **1**(1): 3–15.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

Delaney, L. (2018), 'Behavioural Insights Team: ethical, professional and historical considerations' *Behavioural Public Policy*, **2**(2): 183–189.
Diemand-Yauman, C., D. Oppenheimer and E. Vaughan (2011), 'Fortune favours the (): Effects of disfluency on educational outcomes' *Cognition*, **118**(1): 111–115.
Downs, A. (1957), *An Economic Theory of Democracy*,. Harper & Row Publishers: New York.
Goldhill, O. (2019), '*Politicians love nudge theory. But beware its doppelgänger, "sludge"*'. Quartz. [Online] [Date accessed: 27/07/2020]: https://qz.com/1679102/sludge-takes-nudge-theory-to-new-manipulative-levels/
Gigerenzer, G. (2015), 'On the Supposed Evidence for Libertarian Paternalism' *Review of Philosophy and Psychology*, **6**: 361–383.
Hagman, W., D. Andersson, D. Västfjäll and G. Tinghög (2015), 'Public Views on Policies Involving Nudges' *Review of Philosophy and Psychology*, **6**: 439–453.
Halpern, D. (2015), *Inside the Nudge Unit: How small changes can make a big difference*,. UK: WH Allen.
Hattke, F., D. Hensel and J. Kalucza (2019), 'Emotional Responses to Bureaucratic Red Tape' *Public Administration Review*, **80**(1): 53–63.
Hausman, D. and B. Welch (2010), 'Debate: To Nudge or Not to Nudge' *The Journal of Political Philosophy*, **18**(1): 123–136.
Herd, P., T. De Leire, H. Harvey and D. Moynihan (2013), 'Shifting Administrative Burden to the State: The Case of Medicaid Take-Up' *Public Administration Review*, **73**(1): 69–81.
Hern, A. (2020), '*Twitter aims to limit people sharing articles they have not read*' The Guardian. [Online] [Date accessed: 03/08/2020]: https://www.theguardian.com/technology/2020/jun/11/twitter-aims-to-limit-people-sharing-articles-they-have-not-read
Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and Boosting: Steering or Empowering Good Decisions' *Perspectives on Psychological Science*, **12**(6): 973–986.
Ip, E., A. Saeri, M. Tear (2018), '*Sludge: how corporations "nudge" us into spending more*' The Conversation. [Online] [Date accessed: 27/07/2020]: https://theconversation.com/sludge-how-corporations-nudge-us-into-spending-more-101969
Kahneman, D. (2011), *Thinking, Fast and Slow*, UK: Penguin Books.
Kahneman, D. and A. Tversky (1979), 'Prospect Theory: An Analysis of Decision Under Risk' *Econometrica*, **47**(2): 263–291.
Kroese, F., D. Marchiori and D. de Ridder (2016), 'Nudging healthy food choices: A field experiment at the train station' *Journal of Public Health*, **38**: 1–5.
Lades, L. and L. Delaney (forthcoming), 'Nudge FORGOOD' *Behavioural Public Policy*. DOI: 10.1017/bpp.2019.53
Loewenstein, G., C. Bryce, D. Hagmann, S. Rajpal (2015), 'Warning: You are about to be nudged' *Behavioral Science and Policy*, **1**(1): 35–42.
Madrian, B., D. Shea (2001), 'The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior' *The Quarterly Journal of Economics*, **116**(4): 1149–1187.
Mersch, D. (2018), '*Obfuscated Transparency*' in Alloa, E, Thomä, D (eds.) '*Transparency, Society and Subjectivity*' (2018), SpringerLink
Mill, J. S. (2001), *On Liberty*, (1859). Ontario: Batoche Books Limited.
Mills, C. (2018), 'The Choice Architect's Trilemma' *Res Publica*, **24**: 395–414.
Mills, S. (forthcoming), 'Personalized Nudging' *Behavioural Public Policy*. DOI: 10.1017/bpp.2020.7
Motyka, S., R. Suri, D. Grewal and C. Kohli (2016), 'Disfluent vs. fluent price offers: paradoxical role of processing disfluency' *Journal of the Academy of Marketing Science*, **44**(5): 627–638.
Moynihan, D., P. Herd and H. Harvey (2015), 'Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions' *Journal of Public Administration Research and Theory*, **25**(1): 43–69.

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press

Nobel, N. (2018), '*Nudge vs. sludge – the ethics of behavioral interventions*'. Impactually. [Online] [Date accessed: 27/07/2020]: https://impactually.se/nudge-vs-sludge-the-ethics-of-behavioral-interventions/

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

Oliver, A. (2019), 'Towards a New Political Economy of Behavioral Public Policy' *Public Administration Review*, **79**(6): 917–924.

Rawls, J. (1971), *A Theory of Justice*, Oxford: Oxford University Press.

Rebonato, R. (2014), 'A Critical Assessment of Libertarian Paternalism' *Journal of Consumer Policy*, **37**(3): 357–396.

Reijula, S. and R. Hertwig (forthcoming), 'Self-nudging and the citizen choice architect' *Behavioural Public Policy*, DOI: 10.1017/bpp.2020.5

Sætra, H. (2019), 'The tyranny of perceived opinion: Freedom and information in the era of big data' *Technology in Society*, **59**, p. 101155

Sætra, H. (2020), '*Liberty, Psychological Force and Algorithmic Power: Why liberal political theory also takes issue with technologies of control*' Unpublished Manuscript.

Sanders, M., V. Snijders and M. Hallsworth (2018),, 'Behavioural science and policy: where are we now and where are we going?', *Behavioural Public Policy*, **2**(2): 144–167.

Shahab, S. and L. Lades (2020), '*Sludge and Transaction Costs*' Working Paper 202007, Geary Institute, University College Dublin.

Soman, D. (2020), '*Sludge: A Very Short Introduction*' BEAR. [Online] [Date accessed: 18/09/2020]: https://www.rotman.utoronto.ca/-/media/Files/Programs-and-Areas/BEAR/White-Papers/BEARxBIOrg-Sludge-Introduction.pdf?la=en&hash=DCB98795CB485977A04DDB27EFD800C3DA40220E

Soman, D., J. Xu and A. Cheema (2010), 'Decision points: A theory emerges' *Rotman Magazine Winter*, pp. 64–68

Steffel, M., E. Williams and R. Pogacar (2016), 'Ethically Deployed Defaults: Transparency and Consumer Protection Through Disclosure and Preference Articulation' *Journal of Marketing Research*, **53**: 865–880.

Sunstein, C. (1996), 'Social Norms and Social Roles' *Columbia Law Review*, **96**(4): 903–968.

Sunstein, C. (2012), '*Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych*', SSRN. [Online] [Date accessed: 10/07/2019]: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2171343

Sunstein, C. (2013), 'The Storrs Lectures: Behavioral Economics and Paternalism' *The Yale Law Journal*, **122**(7): 1670–2105.

Sunstein, C. (2019), 'Sludge Ordeals' *Duke Law Journal*, **68**: 1843–1883.

Sunstein, C. (2020), 'Behavioral Welfare Economics' *Journal of Benefits and Costs*, **11**(2): 196–220.

Sunstein, C. (forthcoming), 'Sludge Audits' *Behavioural Public Policy*. DOI: 10.1017/bpp.2019.3.2

Sunstein, C. and J. Gosset (forthcoming), 'Optimal Sludge? The Price of Program Integrity' *Duke Law Journal*. Accessed via SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3642942

Tocqueville, A. (2000), *Democracy in America*, USA: University of Chicago Press.

Thaler, R. (2018), 'Nudge, not sludge' *Science*, **361**(6401): 431

Thaler, R. and C. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth and Happiness*, UK: Penguin Books.

Thunström, L. (2019), 'The welfare effects of nudges' *Judgment and Decision Making*, **14**(1): 11–25.

Tyers, R. (2018), 'Nudging the Jetset to Offset: Voluntary Carbon Offsetting and the Limits to Nudging' *Journal of Sustainable Tourism*, **1**(1): 1–19.

Wendel, S. (2016), '*Behavioral Nudges and Consumer Technology*' in S. Abdukadirov (eds.) *Nudge Theory in Action: Behavioral Design in Policy and Markets*, (2016). Palgrave MacMillan: UK

https://doi.org/10.1017/bpp.2020.61 Published online by Cambridge University Press