

# Salvaging the concept of nudge

Yashar Saghai

## Correspondence to

Yashar Saghai, MA, PhD,  
Berman Institute of Bioethics,  
Johns Hopkins University,  
1809 Ashland Avenue,  
Baltimore, MD 21205, USA;  
ysaghai@jhu.edu

Received 11 April 2012

Revised 5 December 2012

Accepted 17 December 2012

Published Online First

20 February 2013

## ABSTRACT

In recent years, 'nudge' theory has gained increasing attention for the design of population-wide health interventions. The concept of nudge puts a label on efficacious influences that preserve freedom of choice without engaging the influencees' deliberative capacities. Given disagreements over what it takes genuinely to preserve freedom of choice, the question is whether health influences relying on automatic cognitive processes may preserve freedom of choice in a sufficiently robust sense to be serviceable for the moral evaluation of actions and policies. In this article, I offer an argument to this effect, explicating preservation of freedom of choice in terms of choice-set preservation and noncontrol. I also briefly explore the healthcare contexts in which nudges may have priority over more controlling influences.

## INTRODUCTION

Behavioural economist Richard Thaler and law scholar Cass Sunstein (hereafter, T&S) popularised the term 'nudge' in a 2008 book of the same title.<sup>1</sup> Although the authors never provide a technical definition of nudge, we can reconstruct their view. For T&S, A nudges B when A makes it more likely that B will  $\phi$ , by triggering B's automatic cognitive processes, while preserving B's freedom of choice (p. 1–20).<sup>1</sup> The use of nudges has gained particular momentum in public health, health policy and health promotion—where actions and policies can affect individuals' health behaviour and use of the healthcare system—to obtain significant group-level effects. The concept of nudge matters for the moral evaluation of actions and policies because it is designed to pick up efficacious influences that preserve freedom of choice, yet bypass the deliberative capacities of those influenced. The moral import of the concept of nudge thus primarily arises from the conditions requisite for an influence truly to preserve freedom of choice.

Consider a sample of health-affecting influence attempts labelled 'nudge' in the book by T&S or its official online companion edited by John Balz,<sup>1</sup> *The Nudge blog*:

**Asparagus-Lovers.** An investigator suggests to research participants that they liked or loved asparagus during childhood the first time they tried it, creating a false memory, and a false belief about the taste of asparagus. Subsequently, participants report increased general liking of asparagus, greater desire to eat it, and willingness to pay more for it.<sup>2</sup>

**Cafeteria.** A cafeteria manager places healthy food at eye-level at the beginning of the food queue. Unhealthy food comes last and is least visible. The customer is then more likely to purchase healthy food (p. 1–3).<sup>1</sup>

**Deposit Contract.** All primary care physicians of a healthcare system offer their patients the possibility to voluntarily deposit an agreed-upon sum of money with the physicians. The physicians will then return it to the patients in small installments if the latter meet certain agreed-upon objectives to improve their health (e.g., losing weight, exercising, quitting smoking) (p. 232).<sup>1</sup>

**Generic Medication.** Medicare beneficiaries are given generic medication by default but are offered the option of getting the brand-name drug (p. 169).<sup>1</sup>

**HIV-Test Cash Transfer.** In Malawi, residents who pick up their HIV-test results receive 10% of their daily wage in cash.<sup>3</sup>

**Less Than You Think.** University campuses in Montana organize an alcohol consumption-reduction campaign accurately stating that 81% of Montana college students have four or fewer alcoholic drinks each week. The campaign underlines the fact that the majority of students binge-drink less often than what most students assume (p. 68).<sup>1</sup>

**Paternal Competition.** In an Indian village, health professionals post children's medical test results in a public place, creating competition among fathers to improve their children's health.<sup>4</sup>

**Quadruple Bypass Burgers.** The Heart Attack Grill offers free "Quadruple Bypass Burgers" (8,000 calories) for people weighing 350 pounds or more.<sup>5</sup>

Although some of these interventions may unambiguously preserve freedom of choice, sceptics argue that others do not. For them, preserving freedom of choice requires more than avoiding the use of outright coercion, because we also care about 'the control an individual has over his or her evaluations and choices' (p. 128).<sup>6</sup>

Some suspect that nudges may subject us to the control of others because of the mechanisms through which they operate. According to T&S, the mind processes information through two distinct systems (p. 20).<sup>1</sup> 'System 1' is automatic, unconscious, uncontrolled, heuristic, fast and cognitively parsimonious. 'System 2' is reflective, conscious, controlled, analytic, slow and cognitively demanding. If nudges harness the automatic processes of system 1, as T&S tell us, then their influence is efficacious precisely because the influencee does not deliberate over her choice. Why should we not therefore believe that nudges are controlled by nudgers and have no real opportunity to resist an influence attempt easily?<sup>7 8</sup> Unless this question is answered, nudge is nothing but a catchy metaphor.



Editor's choice  
Scan to access more  
free content



- ▶ <http://dx.doi.org/10.1136/medethics-2012-101060>
- ▶ <http://dx.doi.org/10.1136/medethics-2012-101061>
- ▶ <http://dx.doi.org/10.1136/medethics-2012-101109>
- ▶ <http://dx.doi.org/10.1136/medethics-2012-101110>
- ▶ <http://dx.doi.org/10.1136/medethics-2012-101111>
- ▶ <http://dx.doi.org/10.1136/medethics-2012-101112>

**To cite:** Saghai Y. *J Med Ethics* 2013;**39**:487–493.

<sup>1</sup>Balz worked as a researcher on *Nudge* while completing his PhD under Sunstein.

In what sense do nudges preserve freedom of choice? T&S seem to waver between two views: either nudges preserve freedom of choice simply because the nudger does not foreclose any option, or nudges preserve freedom of choice in a more substantive sense. If they defend the former view, many influences will qualify as nudges, and policy-makers will have strong reasons to consider them as prominent regulatory tools. The drawback of this first view is that nudges' moral appeal will be weak for those concerned with the control individuals have over their choices in the face of attempted influence. If T&S choose the latter view, nudges will have enhanced moral traction, but their place in the regulatory toolbox will be more humble because fewer influences will qualify as nudges.

In this article, I argue that influences triggering cognitive processes that bypass deliberative capacities may preserve freedom of choice in a morally robust enough sense, centred on the issue of noncontrol. I thus intend to salvage the concept of nudge from the charge of paying lip service to freedom of choice, but the rescue operation cannot be performed without clarifying T&S's understanding of the concept of nudge. In a concluding section, I explore the healthcare contexts in which nudges may be preferable to more controlling influences.

Let me add a word of caution. This article focuses on the nature and function of nudges, regardless of whether they benefit the recipient, third parties,<sup>ii</sup> or the nudger. Therefore, I will not address worries about paternalism that have attracted much attention in the nudge debate. Many believe nudges and T&S's 'libertarian paternalistic' interventions are co-extensive. This is not the case, although there are overlapping problems. The confusion arises in part owing to T&S defending libertarian paternalism<sup>9</sup> before they popularised the term 'nudge'.<sup>1</sup> The editor of *The Nudge blog* makes the distinction between nudge and libertarian paternalism explicit:

'It's important to point out that nudging complements a libertarian paternalism outlook about public policy, but the two are distinct concepts. Libertarian paternalism is intended as means to help people make decisions that make them better off as defined or judged by themselves—not by a government or private authority. While the nudges cited in the book are intended to do exactly this, nudging takes place in [a] variety of realms where the nudger's explicit goal is to promote [the nudger's] own welfare (think of almost any consumer marketing strategy or retail store layout).'<sup>10</sup>

Libertarian paternalism is a justificatory strategy for a subset of nudges satisfying two conditions: (1) they are performed for the benefit of the nudgee; (2) they satisfy an informed desire welfare criterion. Generic Medication, for instance, does not satisfy the first condition (because it is primarily designed to contain the cost of healthcare), and therefore cannot be supported solely on paternalistic grounds.

## CONDITIONS FOR THE PRESERVATION OF FREEDOM OF CHOICE

To qualify as a nudge, an influence must preserve freedom of choice, but there are numerous conceptions of what freedom of choice entails.<sup>11</sup> The selection of the adequate conception depends on the kind of moral work we want the concept of nudge to do for us. I take it that, in the policy contexts this

concept has been discussed, its moral role is to preempt concerns about interference with choice.

Given this function, at least when we have alternative courses of action open to us, an influence leaves us free to choose. Can we go further and claim that an influence preserves freedom of choice if and only if those influenced have unlimited choice, or even the most extensive set of feasible alternatives? The answer is no. As Alan Wertheimer notes, we 'always choose from among a limited set of options' (p. 10).<sup>12</sup> So long as we interact with one another, the actions of others will affect the range of options that are open to each of us at any time. The choice-set must be preserved in a relative sense:

The Choice-Set Preservation Condition. A preserves B's choice-set when the choice-set is unaltered or expanded, compared to a baseline representing B's situation prior to A's influence attempt (p. 15).<sup>1</sup>

Choice-set preservation is a necessary condition, but not a sufficient condition, for the preservation of freedom of choice. Interference with a person's choices may occur without recourse to coercion. An influence does not sufficiently preserve freedom of choice if we are unable to easily resist it.

T&S endorse The Choice-Set Preservation Condition, and reject the libertarian view that freedom of choice entails maximising 'unfettered liberty of choice' (p. 1162, fn 11).<sup>9</sup> Despite their nominal efforts to accommodate libertarians, T&S do not claim that influences over choices should be minimised, or that only influences that have been expressly consented to preserve freedom of choice. I agree with them. We have no reason to adopt a libertarian standard for the preservation of freedom of choice. We may care for freedom of choice without thinking that every human interaction is an assault on freedom and human dignity unless it has been expressly consented to. Such a standard would rule out the most trivial human informal practices. I maintain in the final section of this article that the libertarian view would fail to protect liberties that matter most, treating all liberties as though they were on a moral par.

T&S gesture towards a condition of noncontrol when they write that, '(t)o count as a mere nudge, an intervention must be easy and cheap to avoid' (p. 6).<sup>1</sup> However, they confess not to have any 'clear definition of "easily avoided"' (p. 248–249).<sup>1</sup> I suspect they are aware that the influencee cannot easily opt out of the arrangements they favour because nudges often unconsciously alter the way we perceive the options we are offered, to make them seem either more attractive or more repulsive. My suggestion is therefore to revise our understanding of what a nudge is by making explicit this second condition for the preservation of freedom of choice that I shall call, following Faden and Beauchamp,<sup>13</sup> 'substantial noncontrol', to guarantee that the influencee can easily resist the influencer.

Influences can be situated on a continuum from fully controlling to fully noncontrolling.<sup>13</sup> Coercion is always controlling, while persuasion is never controlling (as the persuadee willingly accepts the reasons she is given), but there is a third category of influences that do not rely on the means typical of coercion (threats) or persuasion (reasons). There are two thresholds on the continuum with respect to this third category: some influences are substantially controlling, while others are substantially noncontrolling. Faden and Beauchamp define full control, but not substantial noncontrol, which I suggest to characterise as:

The Substantial Noncontrol Condition. A's influence to get B to  $\phi$  is substantially noncontrolling when B could easily not  $\phi$  if she did not want to  $\phi$ .

<sup>ii</sup>Examples include nudges promoting vaccination programmes that do not benefit the person getting immunized.

To summarise, an influence preserves freedom of choice if and only if it preserves the choice-set and is fully or substantially noncontrolling. Critics might reply that, libertarianism to the side, I should also add the preservation of liberty and autonomy to these conditions. There are, however, good reasons for avoiding the introduction of liberty and autonomy to explicate the preservation of freedom of choice.

For one, liberty is a normatively richer concept than freedom of choice. Take MacCallum's characterisation of liberty as a triadic relation: (1) an agent is (2) free from certain constraints, (3) to do or be certain things.<sup>14</sup> A variety of conceptions of liberty can be generated out of this matrix.<sup>iii</sup> Explicating freedom of choice in terms of the preservation of liberty would only generate more confusion. Similarly, there is no consensus about the meaning of autonomy. For instance, autonomy may require not only substantial noncontrol, but also voluntariness, authenticity and any number of additional conditions.

The onus, therefore, is on those who wish to define the preservation of freedom of choice in terms of liberty or autonomy to explain why they are dissatisfied with choice-set preservation and full or substantial noncontrol. Adopting the view I suggest does not commit us to the contentious claim that the preservation of freedom of choice is a sufficient, or even a necessary, condition for the justified use of nudges. Nonetheless, whether an influence preserves the choice-set and is fully or substantially noncontrolling matters for the moral evaluation of actions and policies, if we care about the degree of control exerted by others over at least some of our choices, actions and preferences.

Critics may also object that full or substantial noncontrol is an unhelpful condition because it cannot be readily ascertained. This is why I elaborate on a criterion of easy resistibility to ascertain substantial noncontrol—a concept more difficult to delineate than full noncontrol and more relevant to the analysis of nudges. Here again, Faden and Beauchamp's work on resistibility is a useful starting point. They note that the capacity to resist an influence is subjective: it depends on each person's psychological vulnerabilities. Resistibility is a criterion for testing the degree to which an influence is controlling. For public policy purposes, Faden and Beauchamp defend an 'objective' interpretation of easy resistibility, which depends on the predictable reaction of the 'average (or normal, reasonable, etc.) person' (p. 260).<sup>13</sup> However, the authors rely heavily on the concepts of resistance and resistibility without offering an analysis of these concepts, a lacuna they readily acknowledge (p. 360–361).<sup>13</sup> To supplement their view, I define these concepts and distinguish resistibility from easy resistibility:

**Resistibility.** A's influence is resistible if B is able to oppose the pressure to get her to  $\phi$  if she does not want to  $\phi$ .

**Easy Resistibility.** A's influence is easily resistible if B is able to effortlessly oppose the pressure to get her to  $\phi$  if she does not want to  $\phi$ .

These definitions are consistent with, and inspired by, the psychological literature.<sup>15</sup> What distinguishes resistibility from easy resistibility is the amount of effort the influencee needs to exert in order to oppose the influencer's pressure to make it more likely that she will  $\phi$ . If my explication of freedom of choice is

compelling, the next task is to provide theoretical and empirical evidence to the effect that at least some influences triggering automatic cognitive processes are easily resistible.

## EASILY RESISTIBLE NUDGES

Let me first return to the way nudging mechanisms have been characterised. It is assumed that these influences trigger automatic cognitive processes that always bypass deliberation, but compare two candidate nudges from T&S's sample, for instance, Cafeteria and Less Than You Think. The former example involves no deliberation at all (only unconsciously processed cues), while the latter primarily triggers anchors (a mental shortcut) and social norms to channel deliberation into one predetermined path without entirely bypassing it. In fact, all nudges rely on what I call 'shallow cognitive processes'. I will use the term to cover both nondeliberative and incompletely deliberative processes that share three properties: (1) they are fast; (2) the 'cognitive miser' is inclined to rely on them because they consume few resources; (3) they yield responses that are not the result of full-blown deliberation (ie, the exploration of a broad, if not always exhaustive, range of hypotheses for solving a problem).<sup>16</sup> Replacing the idea of 'automatic cognitive processes' by that of 'shallow cognitive processes' is an important amendment to T&S's account of nudge. This amendment matters because the conditions for easy resistibility vary depending on whether deliberation is entirely or partly bypassed. Whether and how an influence engages our deliberative capacities is also a consideration relevant to the moral evaluation of actions and policies.

To reformulate the question at hand, why should we believe that an influence could trigger shallow cognitive processes and be easily resistible? I suggest analysing the ability to resist an influence easily in the following terms:

Ability for Easy Resistibility. B is able to easily resist A's influence when:

1. B has the capacity to become aware of A's pressure to get her to  $\phi$  (attention-bringing capacities);
2. B has the capacity to inhibit her triggered propensity to  $\phi$  (inhibitory capacities);
3. B is not subject to an influence, or put in circumstances that would significantly undermine the relatively effortless exercise of attention-bringing and inhibitory capacities.

There is sufficient psychological evidence to believe that attention-bringing capacities can be activated even when an influence is 'covert', that is, unannounced, and therefore not explicitly indicated to the influencee. Unconscious preattentive monitoring processes monitor the output of various cognitive processes and alert the conscious attention to anomalies. Preattentive processes recognise a mismatch between the individual's aim (broadly understood) and her behavioural response, calling for the resolution of a conflict.<sup>16</sup> There are various labels for, and models of, these attention-bringing capacities. For instance, according to some cognitive psychologists, 'stimuli that produce a strong feeling of rightness (ie, an intuition that the decision is correct) are not likely to be further scrutinized', whereas those that produce a feeling of dysfluency are more likely to trigger scrutiny (p. 187).<sup>17</sup> At least when individuals have strong and settled enough preferences, goals, or beliefs, they are likely to become aware of an anomaly.<sup>15</sup>

The goal of the influencer is to change the influencee's behaviour or mental states. This change can take three different forms depending on the influencee's disposition before the influence. The influencer can attempt to counter, facilitate, or shape the

<sup>iii</sup>Claims about the compatibility of nudges with freedom of choice are independent from the metaphysical problem of the freedom of the will (see chapter 2 of White).<sup>8</sup>

influencee's behaviour or mental states. Let me explain each possibility in turn.

First, when the influencer's and the influencee's preferences, goals, or beliefs are dissonant, the influencer may attempt to counter them, and steer the influencee in the direction she desires. If there is sufficient conflict between the influencer's and the influencee's aims, the attention-bringing capacities are likely to be activated, and the influencee may be in a position to resist the attempt to counter her behaviour.

Second, when the influencee and the influencer share the same preferences, goals, or beliefs, the influence may facilitate the influencee's action. Facilitation occurs when the influencer removes an obstacle impeding the performance of an action the influencee contemplates. If the obstacle is external to the influencee and the influencee is wholeheartedly committed to  $\phi$ ing, the problem of control does not arise. However, when the obstacle is internal to the agent because her first-order and second-order desires are not aligned (eg, she wants to smoke, but would prefer quitting to run a marathon), facilitation may create the conditions for the activation of attention-bringing capacities because the influencee may, at that moment, not wish to  $\phi$  (eg, her first-order desires are stronger than her second-order desires).

To illustrate the distinction between countering and facilitating, consider Generic Medication. The aim of the policy might be to counter the behaviour of a segment of the target population (those who opt for brand-name drugs for no good reason, increasing the total cost of healthcare), and to facilitate the behaviour of another segment (those who may wish to buy generics but do not know the generic names of the drugs they take).

Third, when the influencee has no preferences, goals, or beliefs before the intervention, the influencer may wish to shape them *de novo*. By this I mean that the influencee's preferences, goals, or beliefs with respect to  $\phi$ ing cannot be settled by derivation from her previous mental states. In that case, the Easy Resistibility criterion does not apply. Remember the counterfactual definition of Easy Resistibility: A's influence is easily resistible if B is effortlessly able to oppose the pressure to get her to  $\phi$  if she does not want to  $\phi$ . Being indifferent between  $\phi$ ing and not  $\phi$ ing cannot give rise to the wish not to  $\phi$ . So these influences are not nudges because they do not qualify as substantially noncontrolling (although they may not be substantially controlling either, but fall somewhere along the spectrum). It is debatable how often health influences are truly *de novo* shaping. Yet we should not hastily conclude that *de novo* shaping influences are morally impermissible just because they might be substantially controlling. Even a fully controlling influence is sometimes morally justified; for example, when coercion is used to avoid harm to others. Similarly, easily resistible countering and facilitating influences are not necessarily morally justified, even if we assume that they do not raise other moral concerns, such as an unattractive risk-benefit profile. The influencer may lack the legitimacy to interact with the influencee on those terms.

In addition to attention-bringing capacities, inhibitory capacities are also necessary for Easy Resistibility. I refer here to our capacity to inhibit our propensity to do what the influencer wants us to do. Inhibition is the capacity to stop a cognitive process once it is triggered,<sup>16</sup> and it is a necessary component of resistance. Some call 'willpower' any type of inhibitory capacity,<sup>18</sup> but these capacities also include the capacity to inhibit erroneous but spontaneous forms of reasoning, for instance.<sup>16</sup>

From a practical perspective, the key condition of Easy Resistibility is (3): it is an exclusion criterion ruling out interventions that are not easily resistible. The influence should not create circumstances that undermine those capacities, or attempt

to steer away an individual who is already in a cognitively strained situation. To qualify as 'relatively' effortless, the influencee should not be expected to have acquired unusually high skills enhancing her attention-bringing or inhibitory capacities.<sup>iv</sup> In the context of influences exercised on groups of individuals, effortlessness is relative to what we can expect of normal individuals (within that group) with limited inhibitory resources that can be depleted.<sup>18</sup> There are accumulating data about circumstances that undermine people's attention-bringing and inhibitory capacities. Typically, the main factors are time pressure, stress (and stress-inducing poverty), submission to perceived authority, fatigue,<sup>15</sup> anxiety, cognitive load and distraction.<sup>16 18 19</sup>

To see how the criterion of Easy Resistibility is serviceable for differentiating substantially controlling influences from substantially noncontrolling ones, consider:

Pharmaceutical TV Ad. In a TV ad for medication, the list of side effects is read in a monotonous voice while displaying images of butterflies and happy people.

This ad triggers nondeliberative shallow cognitive processes by providing the mind with enjoyable and attention-grabbing stimuli irrelevant to the audio disclosure. Its strategy is to confuse the audience through the association of those stimuli with positive affects, rather than the negative affects typically associated with worrisome side effects (see Boush *et al.*, p. 44 and 48, on 'dazzling' tactics).<sup>19</sup> This influence is resistible, but it is not easily resistible. It is not a nudge.

What about influences that involve incomplete deliberation, relying on mental shortcuts to cause the endorsement of a belief, the formation of an intention, or the alteration of a preference? Unlike influences triggering nondeliberative cognitive processes, these influences have informational content. If an influencer weakens our attention-bringing and inhibitory capacities by recourse to deception, concealment, or misrepresentation of material information, the influence is not easily resistible.

For instance, Less Than You Think qualifies as a nudge, but a misleading campaign inflating or spinning the number of students who do not often binge-drink would undermine the students' ability to make their own choice. This is because their capacity to inhibit endorsing the belief they are exposed to is weakened when they expect the authority communicating with them to be truthful. An influence not being easily resistible does not mean that it is not resistible at all. Some individuals detect the misleading or deceitful content of the influence and are in a position to resist it. Recall that when a policy applies to groups of individuals, easy resistibility is assessed for the average person. Whether an information-rich influence is easily resistible depends on how skillful the target population is. Mathematicians are more likely than the rest of us to resist easily an influence attempt based on misleading statistics. In certain circumstances, we may be responsible for increased vigilance based on the realistic expectation that the influencer may use means that weaken our ability easily to resist him or her.<sup>v</sup>

Compare a situation in which an influence activates incomplete deliberation with an attempt to persuade us rationally.<sup>vi</sup> When

<sup>iv</sup>For an overview of the literature on those skills, see Boush *et al.*<sup>19</sup>

<sup>v</sup>I thank Daniel Dennett for bringing this point to my attention.

<sup>vi</sup>I take it that A rationally persuades B when A induces B to willingly believe, or form the intention to  $\phi$ , primarily by presenting her reasons to  $\phi$  (see Faden and Beauchamp,<sup>13</sup> pp. 261–262, and Wertheimer,<sup>12</sup> p. 292).



rational persuasion occurs, the influencee comes to believe or form the intention to  $\phi$  while being in a position to evaluate epistemically the merits of the reasons the persuader offers in support of his or her view. Criteria necessary for epistemic evaluation are complex, and an exploration of those criteria goes beyond the scope of this article. For my purposes, we can simply state that the influencee needs to use a reliable form of theoretical or practical deliberation to reach her conclusion, given her values and preferences. The persuadee is then in a position to evaluate the persuader's claims epistemically. Rational persuasion is therefore fully noncontrolling. We sometimes say that the force of an argument is normatively 'irresistible', but this is different from the claim that the rational persuader exercises an irresistible pressure on us to change our beliefs.

Influences that trigger an incomplete form of deliberation activate mental shortcuts, not formal reasoning. The reliability of these mental shortcuts depends on the quality of the information extractable from the choice environment the influencer controls. As Stanovich puts it, 'a hostile environment for heuristics is one in which there are few cues that are usable by heuristic processes, or there are misleading cues' (p. 21).<sup>16</sup> Asparagus-Lovers is an example of an arranged hostile environment designed to bring about health benefits. The deceived participants are under the substantially controlling influence of the investigator, who has created an environment to misinform participants about their own taste. In contrast to a hostile environment, an influence providing us with accurate information creates a 'benign' environment in which we can reliably trust our fast and frugal shallow cognitive processes to make sufficiently good decisions without the need for slow and resource-consuming full-blown deliberation.

To conclude, I have shown that under some conditions, influences triggering shallow cognitive processes preserve freedom of choice. The nudges are therefore in control of their choices: they have a real opportunity to dissent from the nudger.

## AN AMENDED CONCEPT OF NUDGE

The implication of this argument is an amended definition of a nudge:

Nudge. A nudges B when A makes it more likely that B will  $\phi$ , primarily by triggering B's shallow cognitive processes, while A's influence preserves B's choice-set and is substantially noncontrolling (ie, preserves B's freedom of choice).<sup>vii</sup>

Let me add a few comments to this definition. A nudge increases the *ex ante* probability that some individuals in a group, or an individual over time, will  $\phi$ . It is a probabilistic success term. A nudge is also an intentional action (of the 'choice architect', to use T&S's words) because a person's freedom of choice can be infringed or preserved only by other agents. Although random features of the natural, social, or built environment, or unintentional behaviour of agents, can also trigger our shallow cognitive processes, they are morally and politically irrelevant, and therefore the individual has not been nudged in the sense I use this term.

My account of nudge amends T&S's definition on two points. First, it clarifies what is meant by the preservation of freedom of choice by elaborating on the importance of substantial noncontrol. Second, it introduces a more fine-grained

understanding of nondeliberative and incompletely deliberative nudging mechanisms. Is this proposal revisionist? The response depends on whether these amendments bring our understanding of nudge closer to the function we want this concept to play in our public debates over health influences. Although my proposal reduces the extension of the concept, it clearly favours a more robust understanding of freedom of choice that has real moral traction.

Influences that trigger shallow cognitive processes without satisfying the Substantial Noncontrol Condition are not nudges because they fail to track influences that preserve freedom of choice in a robust enough sense. To facilitate the conversation, I suggest adding to the taxonomy of influences what I shall call 'behavioural prods':

Behavioral Prod. A prods B when A makes it more likely that B will  $\phi$ , primarily by triggering B's shallow cognitive processes, while A's influence preserves B's choice-set but is substantially controlling.

Behavioural prods are not easily resistible. Asparagus-Lovers is a paradigmatic example of a behavioural prod. Because behavioural prods use the same means (ie, default-setting, framing, anchoring, etc.) as nudges and preserve the influencee's choice-set, they are often confused with nudges when commentators narrowly focus on the influencing technique employed, neglecting the degree of control the influence exercises.<sup>viii</sup>

My distinction between nudges and behavioural prods explains why we are reluctant to call many business-oriented influences 'nudges'. Nudging is often not the only way a person or an institution can influence behaviour. More controlling influences are available. In those situations, opting for nudges demonstrates self-restraint, an attitude in tension with the profit-maximising goals of business-oriented practices. The title of a typical sales technique handbook speaks for itself: 'The irresistible offer: How to sell your product or service in 3 seconds or less'. Salespeople are not always successful in crushing resistance, but it would be naive to ignore their goal. Marketers often prefer prodding to nudging.

Critics could object that health promoters use the very same marketing techniques. This is correct, but the focus of my concern is not on a set of techniques.<sup>20</sup> Influences promoting unhealthy behaviour are often more difficult to resist than those spurring consumers to healthy behaviour. I assume for the sake of the argument that we agree on what counts as healthy behaviour. For evolutionary reasons, we are simply more inclined to eat fatty and sugary food and avoid exercise than the opposite. When self-control is at stake, influences reinforcing our natural propensities are more likely to end up substantially controlling our behaviour than those reorienting our impulses and habits or facilitating choices aligned with our second-order desires. Health promoters are more often nudgers than prodders. However, when nudge policies target heterogeneous populations, a health-promoting policy designed to nudge a segment of the population may have an unintended, although predictable, stronger impact on another, psychologically more vulnerable, segment of the population. From a moral standpoint,

<sup>viii</sup>Of course, the reader does not have to adopt the technical terminology I suggest. If some prefer to use the term 'nudge' to refer very broadly to influences activating shallow cognitive processes, they could distinguish controlling from noncontrolling nudges. I use the term 'behavioral prod' to distinguish a mode of influence that should not be confused with what is usually referred to as a 'shove', that is, the whole class of controlling influences (including coercion).

<sup>vii</sup>A and B stand for individuals, institutions, or populations. Although A and B are typically distinct entities, they can be the same entity at different times (self-nudge).

**Table 1** Taxonomy of influences

| Type of influence   | Degree of control   | Definition  |
|---------------------|---|---|
| Choice elimination  | Fully controlling   | A preemptively removes the possibility to $\varphi$ from B's choice-set.  |
| Compulsion          | Fully controlling   | A uses physical force to get B to $\varphi$ .   |
| Coercion            | Fully controlling   | A threatens to make B worse off if B refuses to $\varphi$ .   |
| Behavioural prod    | Substantially controlling                                 | A makes it more likely that B will $\varphi$ , primarily by triggering B's shallow cognitive processes, while A's influence preserves B's choice-set but is substantially controlling.    |
| Disincentive        | Substantially controlling or substantially noncontrolling | A increases the probability of getting B not to $\varphi$ , primarily by raising the monetary or nonmonetary cost of $\varphi$ ing.   |
| Incentive           | Substantially controlling or substantially noncontrolling | A increases the probability of getting B to $\varphi$ , primarily by providing B with some monetary or nonmonetary benefit.   |
| Nudge               | Substantially noncontrolling                              | A makes it more likely that B will $\varphi$ , primarily by triggering B's shallow cognitive processes, while A's influence preserves B's choice-set and is substantially noncontrolling. |
| Rational persuasion | Fully noncontrolling                                      | A induces B to believe willingly, or form the intention to, $\varphi$ , primarily by presenting her reasons to $\varphi$ .  |

health promoters are responsible for the overall reasonably predictable effects of their policies.

In sum, many influences are distinct from nudges when one varies three parameters:

1. The degree to which the influencers control influencees' behaviour;
2. The range of options influencers make available to influencees;
3. The primary means the influencers use.

Combining these parameters, table 1 summarises some of the most salient types of influences that matter for the moral evaluation of actions and policies.<sup>ix</sup>

According to this taxonomy of influences,<sup>x</sup> the concept of nudge is sufficiently precise to be distinguishable from other types of influences.<sup>xi</sup> It is therefore not excessively vague.<sup>20 24</sup> With these distinctions in hand, we are in a better position to see what is wrong with the sample of influences that I mentioned in the introduction. Asparagus-Lovers is not a nudge because it is deceptive, and therefore not easily resistible. Deposit Contract is a self-imposed disincentive deliberately selected not to be easily resistible. Quadruple Bypass Burgers

and Paternal Competition are designed to arouse strong affective responses and the power of social norms (conformity or male competition) in order to control the influencee substantially. From T&S's original list, only Cafeteria and Generic Medication<sup>xii</sup> qualify as nudges.

Although we have gained in precision, there is room for ambiguous or borderline cases. First, we may wonder whether an influence plays a primary or secondary role in explaining its effect. I have noted that nudges 'primarily' trigger shallow cognitive processes: that is, these processes are always triggered and play a major explanatory role with respect to the effect produced by the influence. Secondary processes may also be triggered and play a minor explanatory role. Occasionally, it is difficult to disentangle the effects of several processes, and therefore to distinguish a nudge from a persuasive influence or an incentive. Second, it is sometimes disputable whether an influence is an incentive, a nudge, or a prod. Consider a clear-cut incentive:

**Gym Bonus.** A company offers bonus salary (\$5,000 a year) to employees if they go to the gym on a regular basis.

Why is this different from HIV-Test Cash Transfer? Both cases have a common point: a recipient is offered some benefit for accomplishing a predetermined behaviour (behavioural conditionality), but in Gym Bonus, the benefit is consequential even for a wealthy professional. An incentive may introduce new reasons for action to motivate behaviour change. HIV-Test Cash Transfer perhaps is not an incentive because its effect is not primarily explained by the magnitude of the benefit it provides. If it occurs just once, receiving 10% of one's daily wage in cash may not be significant enough to explain its effect fully, even for disadvantaged populations. The best explanation for the influence's efficacy is that it triggers the tendency to overvalue proximal and immediate benefits. Is HIV-Test Cash Transfer a nudge, an incentive, or a behavioural prod? The answer depends on how misleading the overvaluation of the benefit is.

## CONCLUSION: THE MORAL IMPORT OF THE CONCEPT OF NUDGE IN HEALTHCARE

My account of nudge emphasises two considerations that matter for the moral evaluation of actions and policies: the degree to which others control our choices and engage our deliberative

<sup>ix</sup>I offer these definitions only as a starting point for discussion.<sup>12 13 21–23</sup> Note that not all forms of speech-based influences are reducible to rational persuasion (eg, deception, flattery, etc.). Incentives may be distinguished from rewards and disincentives from punishments by introducing considerations of merit or desert (see Grant,<sup>22</sup> p. 70).

<sup>x</sup>This taxonomy is not an 'intervention ladder'. These ladders classify interventions according to their degree of restrictiveness, assuming that more restrictive interventions have a heavier burden of justification than less restrictive ones.<sup>21</sup> The taxonomy of degrees of control neither affirms nor denies that the burden of justification increases with the degree of control.

<sup>xi</sup>Grant levels two objections against typologies that do not sufficiently distinguish forms of control (or power, to use her terms) from degrees of control (or power) (see Grant,<sup>22</sup> p. 68). Her first objection is that persuasion should not be viewed as necessarily weaker than coercion, because deception might be more controlling than a threat to use force that lacks credibility. This objection does not undermine my argument, because my focus is on *rational* persuasion, which is certainly less controlling than coercion. In addition, I conceive of all categories in my typology as success terms, whereas Grant confuses successful coercion with attempted coercion. Her second objection is that incentives (or 'bargaining', in her terms) cannot be assigned a particular degree of control falling squarely between coercion and rational persuasion. I agree with Grant on this point, and I maintain that incentives and disincentives are sometimes substantially controlling and sometimes substantially noncontrolling.

<sup>xii</sup>For the sake of the argument, I suppose that there is no penalty attached to opting for a brand name.

capacities. I do not, however, beg the question of the moral permissibility of nudges by claiming that an influence has to be substantially noncontrolling to be morally permissible, or even that less controlling influences are always preferable to more controlling influences, by virtue of the principle of the least restrictive alternative. Policy-makers and proponents of rival moral views may use this amended concept of nudge without having to endorse questionable moral assumptions.

The ethics of nudging is particularly intricate when powerful public or private institutions alter minute aspects of the choice environment to generate population-wide health gains. This article focuses on conditions under which one has control over particular choices, no matter how consequential they are. This is why both those who endorse and those who reject the view that all liberties are on a moral par, and deserve the same presumption in their favour, can use my account of the concept of nudge in their debates over the ethics of nudging.

I can only sketch here the implications for the ethics of nudging of the rejection of a general presumption in favour of liberty, a view I have developed in conjunction with Madison Powers and Ruth Faden.<sup>25</sup> The contribution of particular liberties to leading a self-determining life is the criterion we use to assign liberties different presumptive weight when a state policy interferes with them for the sake of public welfare protection. What we mean here by a self-determining life is a life that is, in its main contours, free from the exercise of power by other individuals and by social and political institutions. It is also a life endowed with enough material resources and opportunities to put the individual in a position actually to elaborate and execute a 'plan of life'. The overarching ethical goal is to secure a substantial degree of control over the broad shape of one's life.<sup>26</sup>

Applying this framework to the ethics of nudging, I contend that some health-affecting choices are so fundamental for leading a self-determining life that they ought to be as fully noncontrolled by others as possible. They ought to benefit from a strong presumption in their favour. The substantial noncontrol that nudges guarantee may sometimes not be sufficiently protective of those liberties (eg, certain end-of-life and reproductive choices).

Many other health-affecting choices are negligible for leading a self-determining life (eg, certain quotidian choices typically interfered with by public health policies). They do not require any presumption in their favour. In that case, even influences more controlling than nudges are, *ceteris paribus*, permissible. The principle of the least restrictive alternative is invalid when public authorities, pursuing a legitimate public health goal, interfere with liberties that have no presumptive weight. Nudges are not always preferable to more controlling influences.

I suspect that it is only when the liberties interfered with are sufficiently weighty to require a presumption in their favour, but do not deserve to be strongly shielded from external interference, that substantially noncontrolling influences are, *ceteris paribus*, permissible (even if less restrictive influences are also feasible) and preferable to more controlling influences. The challenge is then to explain whether and why we should prefer nudges to equally noncontrolling incentives and disincentives. If successful, this line of enquiry will lead us to rethink the scope of application of the principle of the least restrictive alternative, which is often assumed in public health and health policy to justify a systematic preference in favour of nudges.

**Acknowledgements** The author warmly thanks Maggie Little and Madison Powers for their detailed comments on earlier drafts of this article, as well as Daniel Andler, Luc Bovens, Kalle Grill, Edmund Pellegrino, Peter Ubel, Bob Veatch, Alan Wertheimer and anonymous referees. The author has benefited from conversations with David Armstrong, Richard Ashcroft, Bennett Foddy, Marianne Promberger, Gina Rini and Simon Rippon. An earlier version of this paper was given in the James Martin Advanced Research Seminar, Future of Humanity Institute (Oxford University) in May 2011.

**Funding** The Edmund Pellegrino Fellowship provided financial assistance and a 1-month mentorship with Dr Pellegrino. The Cosmos Foundation provided a travel grant that supported the author's visit to the UK in May–June 2011.

**Competing interests** None.

**Provenance and peer review** Commissioned; internally peer reviewed.

## REFERENCES

- 1 Thaler RH, Sunstein CR. *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press, 2008.
- 2 More on the "fool your kids" parenting strategy. In: Balz J, ed. *The Nudge [blog]*. Published 14 May 2009. <http://nudges.org/2009/05/14/more-on-the-fool-your-kids-parenting-strategy/> (accessed 4 Apr 2012).
- 3 The behavioral backdrop to financial incentives. In: Balz J, ed. *The Nudge [blog]*. Published 16 Mar 2011. <http://nudges.org/2011/03/16/the-behavioral-backdrop-to-financial-incentives/> (accessed 4 Apr 2012).
- 4 A little healthy parental competition can lead to healthier kids. In: Balz J, ed. *The Nudge [blog]*. Published 27 Jan 2009. <http://nudges.org/a-little-healthy-parental-competition-can-lead-to-healthier-kids/> (accessed 4 Apr 2012).
- 5 Assorted links. In: Balz J, ed. *The Nudge [blog]*. Published 2 Jun 2010. <http://nudges.org/2010/06/02/assorted-links-33/> (accessed 4 Apr 2012).
- 6 Hausman DM, Welch B. To nudge or not to nudge. *J Polit Philos* 2010;18:123–36.
- 7 Anderson J. Review of Nudge: Improving decisions about health, wealth, and happiness, by Thaler RH, Sunstein CR. *Econ Philos* 2010;26:369–406.
- 8 White MD. *Kantian ethics and economics: autonomy, dignity, and character*. Stanford, CA: Stanford University Press, 2011.
- 9 Sunstein CR, Thaler RH. Libertarian paternalism is not an oxymoron. *Univ Chic Law Rev* 2003;70:1159–202.
- 10 A nudge on a hot button issue: abortion. In: Balz J, ed. *The Nudge [blog]*. Published 1 May 2008. <http://nudges.org/a-nudge-on-a-hot-button-issue-abortion/> (accessed 4 Apr 2012).
- 11 Carter I. Choice, freedom, and freedom of choice. *Soc Choice Welfare* 2004;22:61–81.
- 12 Wertheimer A. *Coercion*. Princeton: Princeton University Press, 1987.
- 13 Faden RR, Beauchamp TL. *A history and theory of informed consent*. New York: Oxford University Press, 1986.
- 14 MacCallum GC Jr. Negative and positive freedom. *Philos Rev* 1967;76:312–34.
- 15 Knowles ES, Linn JA. *Resistance and persuasion*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2004.
- 16 Stanovich KE. *Rationality and the reflective mind*. New York: Oxford University Press, 2011.
- 17 Thompson VA. Dual-process theories: a metacognitive perspective. In: Evans JS, Frankish K. *In two minds: dual processes and beyond*. New York: Oxford University Press, 2009: 171–96.
- 18 Baumeister RF, Tierney J. *Willpower: rediscovering the greatest human strength*. New York: The Penguin Press, 2011.
- 19 Boush DM, Friestad M, Wright P. *Deception in the marketplace: the psychology of deceptive persuasion and consumer self-protection*. New York: Routledge, 2009.
- 20 Bonell C, McKee M, Fletcher A, et al. One nudge forward, two steps back: why nudging might make for muddled public health and wasted resources. *BMJ* 2011;342:241–2.
- 21 Nuffield Council on Bioethics. *Public health: ethical issues*. London: Nuffield Council of Bioethics, 2007.
- 22 Grant RW. *Strings attached: untangling the ethics of incentives*. New York: Russell Sage Foundation, 2011.
- 23 Blumenthal-Barby JS, Burroughs H. Seeking better health care outcomes: the ethics of using the 'nudge'. *Am J Bioeth* 2012;12:1–10.
- 24 Marteau TM, Ogilvie D, Roland M, et al. Judging nudging: can nudging improve population health? *BMJ* 2011;342:263–5.
- 25 Powers M, Faden R, Saghai Y. Liberty, Mill, and the framework of public health ethics. *Public Health Ethics* 2012;5:6–15.
- 26 Saghai M. The concept of nudge and its moral significance: a reply to Ashcroft, Bovens, Dworkin, Welch and Wertheimer. *J Med Ethics* 2013;39:499–501.