

The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?

Pelle Guldberg Hansen*

In recent years the concepts of 'nudge' and 'libertarian paternalism' have become popular theoretical as well as practical concepts inside as well as outside academia. But in spite of the widespread interest, confusion reigns as to what exactly is to be regarded as a nudge and how the underlying approach to behaviour change relates to libertarian paternalism. This article sets out to improve the clarity and value of the definition of nudge by reconciling it with its theoretical foundations in behavioural economics. In doing so it not only explicates the relationship between nudges and libertarian paternalism, but also clarifies how nudges relate to incentives and information, and may even be consistent with the removal of certain types of choices. In the end we are left with a revised definition of the concept of nudge that allows for consistently categorising behaviour change interventions as such and that places them relative to libertarian paternalism.

I. Introduction

"... 'Soft Paternalism' would refer to *actions of government that attempt to improve people's welfare by influencing their choices without imposing material costs on those choices... We can understand soft paternalism, thus defined, as including nudges, and I will use the terms interchangeably here.*" (Cass Sunstein *Why Nudge? The Politics of Libertarian Paternalism* 2014, p. 58).

"It is important to recognise that behavioural economics and so-called 'nudges' are distinct. The former is a scientific subdiscipline; the latter is a particular way to apply its findings to policy, which

holds that policy makers should avoid regulations that limit choice (bans, caps, etc.) but can use behavioural science to direct people towards better choices." (*Regulatory Policy and Behavioural Economics*, (Lunn 2014), OECD report)

Since the publication of Thaler and Sunstein's *Nudge: Improving Decisions about Health, Wealth and Happiness*¹ the concepts of *nudge* and their particular version of soft paternalism called '*libertarian paternalism*' have become concepts of increasing interest and debate among public policy makers and academics alike.^{2,3,4,5,6} However, ensuing discussions⁷ concerned with particular applications, the im-

* Behavioral Scientist, Ph.D., Department of Communication, Business & Information Technologies, Roskilde University.

1 Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Revised and Expanded Edition (New York: Penguin Books, 2009).
 2 Sarah Conly, *Against Autonomy: Justifying Coercive Paternalism* (New York: Cambridge University Press, 2012), at p. 29.
 3 Daniel Kahneman, "Foreword", in Eldar Shafir (ed.), *The Behavioral Foundations of Public Policy*, (Princeton NJ: Princeton University Press, 2012), pp. VII et *sqq.*, at p. VIII.
 4 Pete Lunn, *Regulatory Policy and Behavioural Economics* (OECD Publishing, 2014).
 5 Barry Schwartz, "Why not nudge? A Review of Cass Sunstein's *Why Nudge*", 17 April 2014, available on the internet at: <<http://thepsychreport.com/essays-discussion/nudge-review-cass-sunsteins-why-nudge/>> (last accessed on 17 April 2014).

6 Cass R. Sunstein, *Why Nudge?: The Politics of Libertarian Paternalism* (New Haven: Yale University Press, 2014).

7 See e.g. Ryan Calo, "Code, Nudge, or Notice", 99(2) *Iowa Law Review* (2014), pp. 773 et *sqq.*; Henry Farrell and Cosma Shalizi, "Nudge No More", *New Scientist*, 26 November 2011: <http://www.slate.com/articles/health_and_science/new_scientist/2011/11/does_nudge_policy_work_a_critique_of_sunstein_and_thaler_.html> (last accessed on 26 November 2014); Lunn, *Regulatory Policy and Behavioural Economics*, *supra* note 4; Thomas Ploug, Søren Holm and John Brodersen, "To nudge or not to nudge: cancer screening programmes and the limits", 66(12) *Journal of Epidemiology and Community Health* (2012), pp. 1193 et *sqq.*; Anthony Randazzo, "The Case Against Libertarian Paternalism", 23 April 2013, available on the internet: <<http://reason.com/archives/2013/04/23/the-case-against-libertarian-paternalism>> (last accessed 26 November 2014); Mark D. White, *The Manipulation of Choice: Ethics and Libertarian Paternalism* (New York: Palgrave Macmillan, 2013); Mark D. White, "The richness of personal interests: A neglected aspect of the nudge debate", 23

plications and the acceptability of nudges in public policy and their relationship to libertarian paternalism have paid very little attention to a series of disagreements and ambiguities with regard to the way that the concept of nudge has been defined and how it relates to that of libertarian paternalism.

The confusion exists even amongst researchers. For instance, nudges and libertarian paternalism are often discussed as synonymous, as indicated by the two quotes above.⁸ At the same time it is said that supermarkets, restaurants, and cafeterias are nudging consumers all the time whether they recognise this or not.⁹ Yet it seems that the choice architects of supermarkets, restaurants and cafeterias can hardly be said to be libertarian paternalists (whether they recognise this or not, or whether the references to them as choice architects who nudges are intended as such). So *are* nudges always acts of libertarian paternalism, and are libertarian paternalists *always* nudging by definition? Researchers seem to be confused.

An extension of this confusion is that the application and acceptability of nudges are usually discussed without determining or being sufficiently clear about whether certain requirements apply to the motives of the choice architect for an intervention to count as a nudge.¹⁰ That is, does a nudge have to be rooted in the mind-set of libertarian paternalism? Is it only truly a nudge if applied with this mind-set? That is, is it true, as it is often held by critics, that any proponent of nudge interventions is also necessarily suggesting the doctrine that policy makers should avoid regulations that limit choice?

In addition, a series of conceptual specifics are underdetermined in the current literature. For instance, it is often unclear what the exact relation between nudges and incentives is.¹¹ According to Sunstein and Thaler it depends on whether the material or “cognitive” costs are low enough to make a nudge easy or cheap to avoid.¹² Yet, as illustrated by discussions such as in Mongin and Cozic 2014¹³, that easily becomes a little too vague to be useful in any practical discussion. For instance, while imposing a tax is said not to be a nudge¹⁴, and the same goes for placing candy in an obscure place in the supermarket¹⁵, choosing a charm price or asking costumers to pay 5 cents for plastic bags both count as nudges.¹⁶ But, as someone with philosophical inclinations might ask, where is the objective point of difference to be found between the nudge provided by a 5-cent

tax on plastic bags or placing candy at eye height, and a non-nudge of a 5-dollar tax on plastic bags or placing candy behind the counter? Is there a strict line between nudges and other interventions, is it a continuum, or can an intervention be both a nudge and not a nudge at the same time (especially since cognitive costs seem relative to the individual cognitive capabilities of those nudged)? All of this is currently not clear from the literature.

Besides the confusion on incentives, it also remains in the shadows whether and under what conditions the addition or removal of choice options may count as a nudge and if so whether it should also be regarded as a policy based on libertarian paternalism.¹⁷ On the one hand, the definition of nudge provided by Thaler and Sunstein does not exclude adding choices. On the other, the removal of choice options barred by the definition may, as we shall see, in some cases seem to qualify as nudges. This vagueness and its consequences became clearly illustrated

October 2013, available on the internet at: <<http://blogs.lse.ac.uk/politicsandpolicy/the-richness-of-personal-interests-a-neglected-aspect-of-the-nudge-debate/>> (last accessed 26 November 2014); Paula Zoido-Oses, “The problem with nudge policies is that threaten our freedom to choose to act well”, 9 July 2014, available on the internet at: <<http://blogs.lse.ac.uk/politicsandpolicy/the-problem-with-nudge-policies-freedom-to-choose/#Author>> (last accessed on 26 November 2014).

- 8 See also Calo, “Code, Nudge, or Notice”, *supra* note 7, at p. 773, 775, 783, 785, 786, and 795; Conly, *Against Autonomy*, *supra* note 2, at p. 29-31; Lunn, *Regulatory Policy and Behavioural Economics*, *supra* note 4; Farrell and Shalizi, “Nudge No More”, *supra* note 7; Ploug, Holm and Brodersen, “To nudge or not to nudge”, *supra* note 7; Sunstein, *Why Nudge?*, *supra* note 6, at p. 58.
- 9 See e.g. Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 3; and Alberto Salazar, “Libertarian Paternalism and the Dangers of Nudging Consumers”, 23(1) *King's Law Journal* (2012), pp. 51 *et seq.*
- 10 See e.g. Salazar, “Libertarian Paternalism and the Dangers of Nudging Consumers”, *supra* note 9; and Pierre Schlag, “Nudge, Choice Architecture, and Libertarian Paternalism”, 108(6) *Michigan Law Review* (2010), pp. 913 *et seq.*
- 11 See e.g. the debate concerning user financial incentives as nudges around Adam Oliver, “A nudge too far? A nudge at all? On paying people to be healthy”, 12(4) *Healthcare Papers* (2012), pp. 8 *et seq.*
- 12 Sunstein and Thaler, *Nudge*, *supra* note 1, at p. 8 *footnote*.
- 13 Philippe Mongin and Cozic Mikaël, “Rethinking Nudges”, (HEC Paris Research Paper No. ECO/SCD-2014-1067, 2014).
- 14 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 8.
- 15 Thaler and Sunstein, *Nudge*, (*Ibid*).
- 16 Sunstein, *Why Nudge*, *supra* note 6, at p. 64-65.
- 17 Mongin and Cozic, “Rethinking Nudges”, *supra* note 13, at p. 6 and Schlag, “Nudge, Choice Architecture, and Libertarian Paternalism”, *supra* note 10, at p. 917.

by the debate caused by former New York City Mayor Michael Bloomberg's Big Gulp Ban.¹⁸ While Mayor Bloomberg referred to it as a nudge, Richard Thaler quickly tweeted "To state the obvious: a BAN is not a NUDGE. The opposite in fact. So don't blame Bloomberg's ban on large soda cups on us."¹⁹ However, commentators were not convinced on theoretical grounds by Thaler's disclaimer.²⁰

Finally, as we shall see, confusion is widespread even as to how nudges may be properly separated, if at all, from trivial standard measures such as the mere provision of information and rational persuasion.²¹ Such trivial interventions are obviously true measures of libertarian paternalism and qualify as nudges on the original definition presented by Thaler and Sunstein. But if they are nudges, then the question arises: what is new about nudging, if anything at all?

Although these disagreements and ambiguities may seem 'just theoretical' they actually pose serious problems for the on-going efforts to apply behavioural science to public policy and other pro-social domains. Without clear and consistent foundational concepts the new policy-paradigm of applied behavioural science may easily come to seem ill founded, leaving the concept of nudge as well as the ideology

of libertarian paternalism vulnerable to accusations of slippery-slopes, claims of conceptual inconsistency, and warnings that nudges may quickly turn into shoves and so forth. It also renders current efforts vulnerable to a misuse or the dilution of the underlying ideas. Any kind of intervention may easily seem to qualify as a nudge, making any practitioner claim that there is little new to the nudge approach and that he or she has always been nudging; and any objection pertaining to libertarian paternalism may seem to concern the use of nudges as well, and *vice versa*.

However, the prevailing confusion about the nudge concept and its relation to libertarian paternalism is quite understandable. As pointed out by Hausman and Welch,²² the explicit definition of nudge provided by Thaler and Sunstein in *Nudge* only provides two *negative* conditions (that is conditions saying what nudges are not), a couple of heuristics for determining what counts as such, and a vast series of examples distributed throughout the book.²³

Yet, to the person well versed in behavioural science, one of these heuristics provides a fundamental theoretical principle for defining nudges based upon the discipline of behavioural economics. This is the heuristic saying that "... a nudge is any factor that significantly alters the behavior of Humans, even though it would be ignored by Econs."²⁴ In their paper "Debate: To Nudge or Not to Nudge" Hausman and Welch at one point revise Thaler and Sunstein's definition of nudge for reasons of consistency with this principle so as to exclude incentives broadly conceived. In their paper "Nudge and the Manipulation of Choice" Hansen and Jespersen revise the definition even further so as to separate the concept of nudges from mere accidental influences in order to accommodate ethically relevant considerations, and in this process they note the principle as a foundational one as well.²⁵ In the paper "Rethinking Nudges" Mongin and Cozic semantically distinguishes this heuristic as a concept of nudge separate from that provided in the original definition.²⁶

In this paper I argue, based on this principle, for revising Thaler and Sunstein's original definition of 'a nudge' on a series of important points. That is, distinct from Mongin and Cozic I do not treat it as a separate concept of nudge, but as the primary sense of it. The aim is to arrive at a more viable definition to guide the discussion of applications of nudges as well

18 Michael M. Grynbaum, "Health panel approves the restriction on sale of large sugary drinks", *New York Times*, 13 September 2012.

19 Richard Thaler, Tweet on Twitter, 31 May 2012, available on the internet: <https://twitter.com/R_Thaler/status/208273339507150849> (last accessed on 27 December 2014).

20 See e.g. Oliver Burkeman, "How Bloomberg's soda ban is a classic example of 'choice architecture'", *Blog on The Guardian*, 10 July 2012, available on the internet at: <<http://www.theguardian.com/commentisfree/oliver-burkemans-blog/2012/jul/10/bloomberg-soda-ban-new-york-freedom>> (last accessed 26 December 2014) and Pelle Guldborg Hansen, "The 'Big Gulp Ban' – a nudge or not?", 8 October 2012, available on the internet: <<http://inudgeyou.com/the-big-gulp-ban-a-nudge-or-not/>> (last accessed 26 December, 2014).

21 Mongin and Cozic, "Rethinking Nudges", *supra* note 13; and Sunstein, *Why Nudge*, *supra* note 6, at p. 58.

22 Daniel Hausman and Brynn Welch, "Debate: To Nudge or Not to Nudge", 18 *Journal of Political Philosophy* (2010), pp. 123 *et seq.*

23 See Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 6.

24 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 8.

25 Pelle Guldborg Hansen and Andreas Maaløe Jespersen, "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy", 1 *European Journal of Risk Regulation* (2013), pp. 3 *et seq.*, at p. 6.

26 Mongin and Cozic, "Rethinking Nudges", *supra* note 13.

as to explicate the relationship between the concept of nudge and that of libertarian paternalism as understood by Thaler and Sunstein. In the process I clarify how nudges relate to and differ from other interventions such as the provision of factual information and rational persuasion as well as how nudges may incorporate incentives while even being consistent with the removal of certain types of choice. Ultimately I put forward the following definition of what a nudge is:

A nudge is a function of (I) any attempt at influencing people's judgment, choice or behaviour in a predictable way (1) that is made possible because of cognitive boundaries, biases, routines and habits in individual and social decision-making posing barriers for people to perform rationally in their own declared self-interests and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts.

I also conclude that in so far as a nudge serves the declared self-interests of those being nudged, it may further be referred to as *libertarian paternalism* since the revised definition of nudge implies that people's behaviour is influenced in ways that work independently of (i) forbidding or adding any rationally relevant choice options, or (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic incentives and so forth. In addition, this definition also implies that libertarian paternalism goes beyond nudging since it follows from it that nudges (iii) work independently of the provision of factual information and rational argumentation, that fall squarely within libertarian paternalism.

II. The Two Concepts

1. Libertarian Paternalism

Richard Thaler and Cass Sunstein originally introduced the concept of 'libertarian paternalism' in their 2003 essay of the same name published in *The American Economic Review*.²⁷ Here they defined a policy as 'paternalistic' "if it is selected with the goal of influencing the choices of affected parties in a way that will make those parties better off"²⁸ where they intend by "better off" that this be "measured as objec-

tively as possible"²⁹ (and not always equating revealed preference with welfare). According to Thaler and Sunstein, while many economists believe the term paternalistic to be derogatory because they think paternalism always involves some kind of coercion, this is not necessarily the case.³⁰ Policies may be selected with the goal of influencing the choices of affected parties in a way that will make those parties better off, but where there is no coercion involved.³¹ They refer to this kind of paternalism as *libertarian paternalism* and ultimately define it as "... an approach that preserves freedom of choice but authorizes both private and public institutions to steer people in directions that will promote their welfare."³²

According to Thaler and Sunstein an approach like that of libertarian paternalism "should be acceptable to even the most ardent libertarian".³³ Of course, many critics have pointed out that Thaler and Sunstein's notion of libertarian paternalism is neither truly 'libertarian', nor truly 'paternalistic',³⁴ and that it is a contradiction in terms³⁵. However, those discussions will not concern us here, as it is the concept of nudge and its relation to libertarian paternalism that are in focus.

In their best-selling book *Nudge* the notion of libertarian paternalism is further refined. It is described as a "movement" or "strategy" recapturing common sense from dogmatists.³⁶ The *libertarian* aspect of the strategy is said to lie in "the straightforward insistence that, in general, people should be free to do what they like – and to opt out of undesirable arrangements if they want to do so".³⁷ Borrowing a

27 Richard H. Thaler and Cass R. Sunstein, "Libertarian Paternalism", 93(2) *American Economic Review* (2003), pp. 175 et seq.

28 Thaler and Sunstein, "Libertarian Paternalism", *supra* note 27, at p. 175.

29 *Ibid.*

30 *Ibid.*

31 *Ibid.*

32 *Ibid.*, at p. 179.

33 *Ibid.*, at p. 175.

34 See e.g. Hausman and Welch, "Debate: To Nudge or Not to Nudge", *supra* note 22.

35 Gregory Mitchell, "Libertarian paternalism is an oxymoron", 99 *Northwestern University Law Review* (2005), pp. 1245, et seq.

36 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 5.

37 *Ibid.*

term from Friedman, Thaler and Sunstein say that “libertarian paternalists urge that people should be ‘free to choose’” and strive to “design policies that maintain or increase freedom of choice.”³⁸ In particular, Thaler and Sunstein say that by modifying the term *paternalism* with *libertarian* they simply mean *liberty preserving*, adding that “Libertarian paternalists want to make it easy for people to go their own way; they do not want to burden those who want to exercise their freedom.”³⁹ According to Thaler and Sunstein the *paternalistic* aspect lies in the claim that “it is legitimate for choice architects to try to influence people’s behaviour in order to make their lives longer, healthier, and better.”⁴⁰ Hence they argue for a self-conscious effort by institutions “to steer people’s choices in directions that will improve their lives.”⁴¹ However, they also modify their understanding of paternalism compared with their 2003 paper, now - referring to Van De Veer⁴² – holding that, “a policy is ‘paternalistic’ if it tries to influence choices in a way that will make choosers better off, *as judged by themselves*”⁴³, rather than interpreting “better off” as earlier to be “measured as objectively as possible”⁴⁴.

According to Thaler and Sunstein “libertarian paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened.”⁴⁵ Yet, it does count as paternalism, or ‘soft paternalism’ because “private and public choice architects are not merely trying to track or to implement people’s anticipated choices. Rather, they are self-consciously attempting to move people in directions that will make their lives

better.”⁴⁶ In particular, Thaler and Sunstein say that by doing this “They nudge.”⁴⁷

2. Nudge

But what do Thaler and Sunstein mean when saying that libertarian paternalist ‘nudge’? The concept was originally suggested by the first editor approached by Thaler and Sunstein, an editor who ultimately declined to publish the book that later became a best-seller titled *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Beforehand most famous from Monty Python’s sketch “nudge, nudge, wink, wink” the concept of ‘nudge’ has now become a term closely tied – if not almost synonymous with – Thaler and Sunstein’s concept of ‘libertarian paternalism’.⁴⁸ But what does ‘nudge’ mean and how does the concept fit together with that of libertarian paternalism?

In *Nudge* Thaler and Sunstein define a nudge as follows:

“A nudge, as we will use the term, is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting fruit at eye level counts as a nudge. Banning junk food does not.”⁴⁹

The concept of nudge thus seems to fit like hand in glove with that of libertarian paternalism. Libertarian paternalism is an approach that authorises both private and public institutions to steer people in directions that will promote their welfare; a nudge alters people’s behaviour in a predictable way. Libertarian paternalism is an approach that preserves freedom of choice; a nudge works without forbidding any options or significantly changing economic incentives. Finally, libertarian paternalists nudge.

But does that mean that nudges are always libertarian paternalistic by definition? As noted in the introduction this has been a tacit assumption by many commentators. It is thus often assumed that interventions based on nudges and the strategy of libertarian paternalism are the same. That is, that nudges are libertarian and paternalistic, and that libertarian paternalistic measures are nudges. But is this the case? In order to answer that question and resolve

38 Ibid.

39 Ibid.

40 Ibid.

41 Ibid.

42 Donald Van De Veer, *Paternalistic Intervention: The Moral Bounds on Benevolence*, (Princeton: Princeton University Press, 1986).

43 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 5.

44 Thaler and Sunstein, “Libertarian Paternalism”, *supra* note 27, at p. 175.

45 Ibid.

46 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 6.

47 Ibid.

48 See *supra* note 8.

49 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 6.

the resulting confusion and ambiguities, we need to look more closely into the concept and the theoretical foundations underpinning it.

III. Nudge by Definition

It is often said that the concept of nudge is ill defined⁵⁰ and may lead to slippery slopes⁵¹. Such observations easily lead to critiques headed “When nudge comes to shove” or something like it.⁵² However, given Thaler and Sunstein’s definition quoted above, such critiques may seem to misfire. In fact, by rearranging the original definition provided by Thaler and Sunstein a bit, it seems that they provide a quite clear-cut definition saying that:

A nudge is any aspect of the choice architecture that alters people’s behaviour in a predictable way without

- 1) forbidding any options or
- 2) significantly changing their economic incentives.

Still, definitions are not all of one kind. They serve a variety of functions. Thus e.g. some definitions are descriptive, while others are stipulative. That is, some definitions try to *describe* or capture the actual usage of the term, while others *stipulate* or impart a meaning to the defined term and thus involve no commitment that the assigned meaning agrees with prior uses (if any) of the term.⁵³ For this latter type, “ill defined” can only mean that a definition fails to provide the wanted conceptual clarity and consistency needed for it to work within its intended area of application.

Obviously, Thaler and Sunstein’s original definition is a *stipulative* one, and one that has since then been broadly adopted. However, as this paper will show, it is not always true for Thaler and Sunstein’s definition of nudge that it provides the intended clarity and consistency. In particular it fails to serve its function relative to its theoretical foundations in behavioural economics and its relation to the concept of libertarian paternalism. Of course, that does not imply that the concept is fundamentally flawed or necessarily will lead to slippery slopes. Rather it implies that it can be improved upon. Throughout this section I thus offer an *explication* of the definition of nudge resulting in what I see as an absolute improvement of an existing, imperfect concept.

1. Econs and Humans

To explicate the definition of ‘nudge’ we first need to understand the purpose intended to be served by its introduction and to which it should be aligned. In *Nudge* Thaler and Sunstein explicitly motivates its introduction with the findings of four decades of behavioural economics.⁵⁴ This is not surprising since Thaler is one of the world’s leading behavioural economists and Cass Sunstein is, amongst many other things, a pioneer and leading scholar in Behavioural Law, a movement explicitly rooted in behavioural economics and the Biases and Heuristics programme of Kahneman and Tversky. This background becomes particularly obvious in Thaler and Sunstein’s section in *Nudge* titled ‘Humans and Econs: Why Nudges Can Help’⁵⁵ that follows directly after their section ‘Libertarian Paternalism’⁵⁶ that ends by providing the original and explicit definition of nudge quoted above.

The distinction between *Econs* and *Humans* introduced by Thaler and Sunstein is a distinction contrasting traditional conceptions of human behaviours as derivatives of axioms of reason and rationality with the insights into human behaviours emerging from cognitive psychology and behavioural economics in the last four decades. Reason and rationality have been analysed most intensely in philosophy

50 See e.g. Mongin and Cozic, “Rethinking Nudges”, *supra* note 13; and House of Commons, *Public Health: Twelfth Report of Session 2010-12, Vol. 1: Report. Together with Formal Minutes* (Great Britain: Parliament: House of Commons: Health Committee 2011), at p. 84.

51 See e.g. Mario J. Rizzo and Douglas Glen Whitman, “Little Brother Is Watching You: New Paternalism on the Slippery Slopes”, 51 *Arizona Law Review* 2009, pp. 685 *et seq*; and Adam C. Smith and Todd J. Zywicki, “Behavior, Paternalism, and Policy: Evaluating Consumer Financial Protection”, George Mason Law & Economics Research Paper No. 14-05 (2014), at p. 12.

52 See e.g. Tim Adams, “Nudge economics: has push come to shove for a fashionable theory?”, *The Guardian*, 1 June 2014, available on the internet at <<http://www.theguardian.com/science/2014/jun/01/nudge-economics-freakonomics-daniel-kahneman-debunked>> (last accessed on 27 December 2014); John Tierney, “A Nudge (or Is it a Shove?) To the Unwise”, *New York Times*, 24 March 2008, available on the internet at <<http://tierneylab.blogs.nytimes.com/2008/03/24/a-nudge-or-is-it-a-shove-to-the-unwise/>> (last accessed on 27 December 2014).

53 See Anil Gupta, “Definitions”, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, (Fall 2014 Edition), available on the internet <<http://plato.stanford.edu/archives/fall2014/entries/definitions/>> last accessed (28 December 2014).

54 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 6-8.

55 *Ibid.*

56 Thaler and Sunstein, *Nudge*, *supra* note 1, p. 4-6.

and economics in a movement that ultimately led to neo-classical axiomatisations of rational or economic behaviour. According to these axiomatisations, behaviour is epitomised by *rational choice* or *decision-making* based upon individual agents reasoning their way by deduction to appropriate actions based on their desires or *preferences* together with their *beliefs* and information about the world. Keeping it simple, options, preferences, and beliefs are thus all that influence the behaviours and actions of a rational agent.⁵⁷

In passing it may be noted that in the neo-classical view of decisions and behaviour there is no room for nudges and no reason for libertarian paternalism. This is perhaps most clear from the adoption of Revealed Preference Theory (RPT) in the neo-classical economic analysis of real world behaviour. RPT assumes that the preferences of consumers are revealed by their purchasing habits. If a consumer chooses a particular item it is inferred that this outcome is observed as a result of the consumer's preference for that item above others given her available information. This means that the intention-action gap that motivates the libertarian paternalist strategy and, as we shall see, within which the nudge approach to behaviour change works, is ruled out *a priori*.

However, as pointed out by Thaler and Sunstein,⁵⁸ during the last four or five decades the disciplines of behavioural economics, cognitive, and social psychology have provided ample evidence or *behavioural insights* that much of our behaviours and choices cannot be explained convincingly as being consistent with the neo-classical axioms of rational behaviour. For instance, behavioural economics has shown

that human behaviour and choice exhibit *bounded rationality*, *bounded self-interest*, and *bounded willpower*.⁵⁹ This helps explain not only why people often fail to reason their way to the right conclusions, but also why we often fail to act upon these conclusions when reached, causing the gap between our good intentions and our actual behaviour - the gap that motivates the doctrine of libertarian paternalism.

Now, to some, "bounded" may seem to carry the connotation that we are merely *limited* in our cognitive skills at exerting rationality, self-interest, and willpower. That is not the case. Behavioural economics runs on experimental evidence pioneered in cognitive and social psychology, showing that much of our individual and social behaviour is due to our brains processing information in ways that are not only bounded but also cognitively biased, where a cognitive bias is a systematic pattern of deviation in judgment or decision-making. Implicit in this concept is a "pattern of deviation" from a baseline or standard of comparison.⁶⁰ In this case the baseline is made up by the bundle of neo-classical assumptions and their derivatives giving rise to the rational agent *Homo Economicus* - or as Thaler and Sunstein dub them in plural: *Econs*.

Of course, the empirical evidence and theoretical constructs of biases invoked to explain our behaviours only make up a consistent alternative to neo-classical economics and its derivatives such as RPT, if an alternative theory exists that not only predicts, but better explains the facts. Here, like others, Thaler and Sunstein rely on *Dual Process Cognitive Theories* (DPTs), especially as portrayed by Kahneman⁶¹ as a result of Kahneman and Tversky's work on the Biases and Heuristics programme.

DPTs assert that the human brain functions in ways that invite a distinction between two kinds of thinking: one, which is intuitive and automatic, and another, which is reflective and rational. Kahneman dubs these ways of thinking System 1 and System 2, respectively; we choose, however, to follow the lead of Thaler and Sunstein when referring to these modes of thinking as automatic thinking and reflective thinking. Automatic thinking is characterised by being fast, intuitive, and usually not associated with experiences that one would describe as thinking. Reflective thinking is associated with the deliberate and conscious processing of information. It is slow, effortful, and dependent on concentration. It is associ-

57 See also Thomas Gilovich and Dale Griffin, "Introduction - Heuristics and Biases: Then and Now", in Thomas Gilovich, Dale Griffin and Daniel Kahneman (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002), pp. 1 *et seq*; Martin Peterson, *An Introduction to Decision Theory*, (Cambridge: Cambridge University Press, 2009).

58 Thaler and Sunstein, Nudge, *supra* note 1, p. 7.

59 See also Christine Jolls, Cass R. Sunstein and Richard Thaler, "A Behavioral Approach to Law and Economics", 50(5) *Stanford Law Review* (1998), pp. 1471 *et seq*.

60 Gilovich and Griffin, "Introduction - Heuristics and Biases", *supra* note 57.

61 See e.g. Daniel Kahneman, "Maps of Bounded Rationality: Psychology for Behavioral Economics", 93(5) *The American Economic Review* (2003), pp. 1449; and Daniel Kahneman, *Thinking, Fast and Slow*, (London: Allen Lane 2011).

Automatic thinking	Reflective thinking
Uncontrolled	Controlled
Effortless	Effortful
Associative	Deductive
Fast	Slow
Unconscious	Self-aware
Skilled	Rule following

Table 1.1

ated with self-awareness, the experience of agency, autonomy, and volition. The key features of each system are shown in Table 1.1.

DPTs explain why we – being *Humans* rather than *Econs* – not only fall short of rational decision-making, but actually may systematically *deviate* from its normative prescriptions (even in our reflected decisions at times) due to our behaviour and decisions being biased by seemingly irrelevant factors and aspects of decision-making and behavioural contexts. At its most general, the concept of nudge is devised to capture the fact that human decision-making and behaviour are influenced by cognitive boundaries and biases in ways that may be utilised for promoting particular behaviours - and Thaler and Sunstein's original definition is their attempt to capture this insight.

2. The Principle from Behavioural Economics

However, the discussion of Human and Econs ultimately leads Thaler and Sunstein to state a corollary of their original definition that I will refer to as *the principle from behavioural economics*. This says:

“In accordance with our definition, a nudge is any factor that significantly alters the behavior of Humans, even though it would be ignored by Econs.”⁶²

This principle of behavioural economics, I believe, although misleadingly stated as a *derivative* of the explicit definition⁶³, is what provides us with the clear-cut and foundational criterion of what should count as a nudge. In accordance with the general gist of Thaler and Sunstein's book, its adoption takes the Biases and Heuristics programme as the core moti-

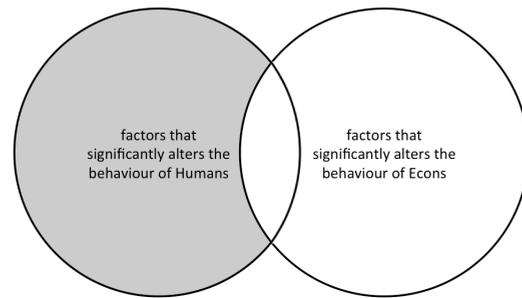


Figure 1

vation for devising a concept of nudge for referring to any contextual feature causing Humans to deviate from the prescriptions of normative rationality. As a criterion it states that if we know what behaviour or choice would count as rational in a given situation provided agents' preferences and information, and we observe judgment, choice, or behaviour deviating from this prediction, then the contextual feature responsible for such deviation should be categorised as a nudge. If accepted as a foundational criterion, it thus follows that any viable definition of a nudge should be aligned with it. See Figure 1.

Granted the criterion as a fundamental one, the concept of *nudge* is thus devised to refer to features that influence behaviour in ways not in accordance with that of economic rationality. Since economic rationality follows from the dictates of principled reason, one could add that nudges work in ways that should not work in principle, or in ways that 'ought' not influence us. That is, in principle a nudge shouldn't influence behaviour, but in practice it does. In particular, it is obvious that condition (1) and (2) of Thaler and Sunstein's original definition are intended to exclude those features that influence the behaviour, not only of humans, but also of rational agents. Hence, it seems that it is conditions (1) and (2) of the original definition that are derivatives of the principle, rather than *vice versa*. Of course, the acceptability of this proposed hypothesis about the foundations of the concept of nudge, rests on the ability of the principle from Behavioural Economics to

62 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 8.

63 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 6.

provide a consistent and clear framework for a more viable definition of the concept than that provided by Thaler and Sunstein's explicit definition. I believe it does, and the remainder of this paper provides the argument for this. However, instead of constructing a definition anew from the foundational principle from Behavioural Economics, I will start in *medias res* and revise the original definition so as to reconcile this with the principle.

IV. Reconciling the Definition of Nudge with the Principle from Behavioural Economics

1. Adding some Positive Conditions

As a first step to reconcile Thaler and Sunstein's definition of nudge with the principle from Behavioural Economics, we may notice that Thaler and Sunstein's original definition of a nudge says nothing about cognitive boundaries and biases (Humans) or the baseline from neo-classical economics (Econs). It only says what a nudge is not. But given what has just been explained about DPTs and the Biases and Heuristics programme, we can now see why Daniel Hausman and Brynn Welch in a later article add the "positive" condition to the definition of a nudge saying that:

"They [nudges] are called for because of flaws in individual decision-making, and work by making use of those flaws".⁶⁴

In fact, being a conjunctive sentence, this addition contains two positive conditions. One of function or *motivation* – that "nudges are called for because of..." and a second condition, which explicates the underlying principle by specifying that nudges "work by making use of those flaws". That is:

(i) They [nudges] are called for because of flaws in individual decision-making, and (ii) work by making use of those flaws.

Thus, as a first step to reconcile the definition of nudge with the principle from Behavioural Economics, we may integrate Hausman and Welch's positive conditions into the former. But rather than calling biases "flaws", I prefer calling them by their proper name: *cognitive bias*. Also, it should be added that biases do not necessarily only pertain to individual decision making, but also may be argued to pertain to *social* decision making, i.e. when aggregate group behaviour deviates from the interests of each individual group member.⁶⁵ In addition, it is not only biases, but also *cognitive boundaries, habits and routines* that may systematically lead Humans to deviate from the behaviour of Econs.⁶⁶ Finally, we should be a bit more specific than Hausman and Welch are, so instead of saying "making use of those flaws", we specify that biases should be an *integral part of the choice architecture*.

Given these adjustments the positive conditions may be added to the original definition so that we now have a definition with four conditions:

A nudge is any aspect of the choice architecture that alters people's behavior in a predictable way without

- (1) forbidding any options or
- (2) significantly changing their economic incentives.
- (3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

2. Incentives (Condition (2))

Still, more work needs to be done to reconcile the definition of nudge with its theoretical underpinnings. As noted above, the fundamental principle for determining if something is a nudge is by determining whether it alters the behaviour of Humans but would be ignored by Econs.⁶⁷ Fundamental to Econs is that their preferences may be captured by a suitable utility-function. As a fundamental construct in economic theory it is not surprising that a utility function is

64 Hausman and Welch, "Debate: To Nudge or Not to Nudge", *supra* note 22, at p. 126.

65 See e.g. Vincent F. Hendricks and Pelle Guldborg Hansen, *Infostorms: How to Take Information Punches and Save Democracy* (New York: Springer, Copernicus, 2014); and Cass R. Sunstein *Infotopia: How Many Minds Produce Knowledge* (New York: Oxford University Press, 2006).

66 See e.g. Judith A. Oullette and Wendy Wood, "Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior", 124(1) *Psychological Bulletin* (1998), pp. 54 *et seq.*

67 Thaler and Sunstein, *Nudge*, *supra* note 1, at p. 9.

sensitive to *economic* incentives – hence, Thaler and Sunstein’s condition (2) saying that a nudge influences behaviour without “significantly changing [people’s] economic incentives”.

Yet, as pointed out by Hausman and Welch,⁶⁸ rational agents are not only responsive to economic incentives. For instance, the utility-function of a rational agent is determined by the prospect of pain as well as penalties. This is not captured by the definition as it stands since, if taken at face value, the definition would render a 10.000 voltage electroshock, or a public beating to count as a ‘nudge.’ Because this would be a rather uncharitable interpretation of what Thaler and Sunstein mean by a nudge, as well as not in accordance with the principle from Behavioural Economics, it is only reasonable to follow Hausman and Welch’s suggestion of broadening the definition so as to encompass all other types of incentives affecting a rational utility function as well.

Thus, we may revise the definition of a nudge as follows:

A nudge is any aspect of the choice architecture that alters people’s behaviour in a predictable way without

- (1) forbidding any options or
- (2) significantly changing **incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth.**
- (3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

a. “Significantly” Changing Incentives

Still, one may ask why the original condition (2) says “significantly” (as well as why Hausman and Welch restates this as “appreciably”). To the best of my judgment this is done to accommodate insights from Behavioural Economics about how changing incentives may be used to influence behaviour of Humans in ways that do not affect Econs.

Let me give three examples of how this may happen.

1. Lotteries. First, disproportionately large behavioural effects on Humans may be obtained by making small, in principle insignificant changes in incen-

tives that wouldn’t have an effect on Econs (at least not to the same extent). Say for instance you want to conduct a survey amongst 500 people. To get people to respond you consider giving them some proper incentive. But your budget is only \$500. One possibility is to offer each potential respondent \$1 to complete the survey. Another possibility is to offer each potential respondent a ticket for a lottery draw where the prize is an iPad. Now, to a rational agent this should make no difference since the expected utility in both scenarios would be \$1. However, for Humans it makes a world of difference. Being offered a lottery ticket increases our tendency to participate since we tend to overestimate small probabilities because of, amongst other factors, the availability heuristic, anchoring on arbitrary priors, and a tendency to be more sensitive to probability changes close to 0 than to probability changes away from 0.⁶⁹ In addition one may also argue that Humans often perceive direct payment differently than they perceive lottery tickets (usually regarding the latter as more like a gift, thereby activating norms of reciprocity, contrary to the former which we regard as impersonal payment, possibly with some suspicion and comparison to alternative activities and the profit gained by the payer).

2. A second example is that of rearranging incentives in ways that shouldn’t have any effect on us if we were Econs, but does affect us, as we are Humans. For instance, to an Econ it shouldn’t affect consumption patterns whether a tax payback is given as a lump sum or distributed in a series over time. However, due to what is called *mental accounting*, there is a tendency for humans to treat a lump sum payment very differently.⁷⁰ For instance, we may regard it as house-money, or decide to use it for larger investments, such as travelling, rather than integrate it into our everyday consumption. In addition, incentives may be distributed over time

68 Hausman and Welch, “Debate: To Nudge or Not to Nudge”, *supra* note 22, at p. 123-136.

69 Zach Burns, Andrew Chiu and George Wu, “Overweighting of small probabilities”, in James J. Cochran (ed), *Wiley Encyclopedia of Operations Research and Management Science* (New York: Wiley, 2010).

70 Shefrin, H. and Richard Thaler, “The behavioral life-cycle hypothesis”, 26 *Economic Inquiry* (1988), pp. 609 *et seq*; Richard H. Thaler, “Mental Accounting Matters”, 12 *Journal of Behavioral Decision Making* (1999), pp. 183 *et seq*.

causing so-called present-bias, where Humans discount valuations hyperbolically rather than exponentially, that is, where valuations fall very rapidly for small delay periods, but then fall slowly for longer delay periods, rather than being discounted at a constant rate.⁷¹

3. A third example is that of “charm prices”, e.g. advertising a price of \$0.99 rather than \$1 thereby utilising the ‘left-digit-right-digit’ effect. This price change shouldn’t lead to the significant changes in consumer behaviour that we actually observe.⁷²

Effects like these three are not reconcilable with the standard economic model of Econs. Yet, all do seem to involve changing incentives. However, looks may be deceiving. Being more precise we may point out that example 1 is not really about *changing* incentives, but rather about *restructuring* incentives in terms of a lottery in ways that would not affect Econs. Likewise example 2 is about *rearranging* incentives, rather than changing them. But what about example 3 – is this not obviously about changing incentives? Well, in a sense changing incentives does create the behavioural effect. Yet, the behavioural effect is not credibly derivable from an agent’s utility function alone. The incentives are not changed significantly in creating a significant effect. The nudge is identified by *the deviation* from the predictions of the standard model from observed behaviour – a deviation caused in this case by the *salience effect* – and not the full behaviour change as such.

It is most likely in order to allow such insights from behavioural economics of how *insignificant* change in incentives may create effects beyond what may be explained as rational that Thaler and Sunstein adopt the qualification “significantly” in the 2nd condition. However, if we substitute “work without” with “that works independently of” we can retain this point without the need of this qualification that has sometimes confused commentators. Also, to emphasise more clearly that it is not the full intervention or choice architecture as such that is evaluated as a nudge – e.g. it is not the incentive as such, but the

structuring of an incentive as a lottery – we may substitute the somewhat vague term ‘aspect’ with that of ‘function’ which is also a more standard term in the behavioural sciences. Hence, we may say:

A nudge is a **function** of the choice architecture that alters people’s behaviour in a predictable way **that works independently of**

- (1) forbidding any options or
- (2) changing their incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth.
- (3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

3. Choices (Condition (1))

So far the only part of the original definition that we have not revised is that concerned with the ruling out or forbidding of options, i.e. condition (1). Obviously, this condition has a strong affinity with the idea of libertarian paternalism. That is, its formulation seems primarily intended to target this complex of ideas, and only secondarily at capturing the underlying principle of a nudge as expressed by the revised definition. In particular, by emphasising that a nudge works without forbidding any options it follows that nudging neither involves bans or mandates, which is a basic characteristic of libertarian paternalism. In practice such a ban or mandate may often be reduced to a positive or negative incentive coupled with a symbolic dimension of social disapproval.

From a conceptual point of view, however, condition (1) turns out to be incomplete along several dimensions. For instance, it does not say anything about whether one may add options to the choice set. This raises the question of whether *adding choices* may qualify as a nudge?

a. Adding Choices: Predictively Rational

To investigate this, assume that Marge wants to influence Homer in a way so as to stop him from eating all that cake. Also assume that in the status quo Homer has the options or choice-set *A* comprising *a*:

71 Edward Cartwright, *Behavioral Economics* (New York: Routledge Advanced Texts in Economics and Finance, 2011), p. 143-147.

72 Manoj Thomas and Vicki Morwitz, “Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition”, 32(1) *Journal of Consumer Research* (2005), pp. 55 *et seq.*

“not eating cake” and b : “eating cake”. Finally we also assume that Homer prefers b to a . That is,

$$A_h = (a, b), a <_h b.$$

Next, assume that Marge in order to influence Homer decides to serve a bunch of sandwiches - i.e., she adds c : “eat sandwiches” to the choice-set - knowing that Homer prefers sandwiches to cake. Thus we have that,

$$A_h = (a, b, c), a <_h b <_h c.$$

The question is whether Marge’s addition of sandwiches to the choice-set should be regarded as a nudge.

Obviously, Marge attempts to influence Homer’s behaviour in a predictable way. Also, she is not forbidding any options. Hence (1) is not violated. Nor is she significantly changing the incentives so (2) is respected as well. Thus, based on the original part of the definition Marge’s addition would seem to qualify as a nudge.

However, it is obvious that Homer’s preferences are rational in the sense that he is consistent in preferring sandwiches to cake. Hence, the intervention would not only influence Humans - or Homer in this case - but also Econs. This is because the influence is not made possible because of cognitive boundaries or bias in individual and social decision-making, or routines or habits, posing a barrier for Homer to perform rationally. Instead, the influence is made possible by Homer’s rational capacities. Also, it does not work by making use of those boundaries and biases as an integral part of Marge’s attempt at creating behaviour change. In conclusion, Marge’s attempt is not a nudge since it does not satisfy conditions (3) and (4) that have been explicated on the basis of the principle from behavioural economics.

b. Adding Choices: Predictively Irrational

But does that mean that we always have to rule out the addition of choice options as a nudge? No, because one may also add choices to the existing choice-set so as to influence people in a “predictively irrational” way. Various decoy effects violating the independence of irrelevant alternatives axiom of decision theory reveal this. Decoy effects are one family of biases in which the asymmetric dominance effect and the compromise effect resulting from extremeness aversion have been shown to be among the most stable.⁷³

To illustrate the asymmetric dominance effect, imagine that you are faced with the choice between

two laptops that vary in price and storage:

Choice-set 1

	A	B
Price	\$400	\$300
Storage	300 GB	200GB

Given this choice set, you may either prefer A for its greater storage capacity or B for its lower price. Now, suppose that a new laptop, C , is added to the choice set, where C is more expensive than both A and B and has more storage capacity than B , but less than A :

Choice-set 2

	A	B	C
Price	\$400	\$300	\$450
Storage	300 GB	200 GB	250 GB

If you are like most other people, the addition of C to the choice-set will make you become more attracted to A . This is because, while you are likely to avoid C since you can get laptop A with more storage for a lower price, C affects your preferences by acting as a basis of comparison for A and B . Because A is better than C in both respects, while B is only partially better than C , more consumers will prefer A now than did before. C is therefore a decoy whose sole purpose is to increase sales of A . This is what is called the asymmetric dominance effect.

The asymmetric dominance effect can also be used to influence your preferences so that you come to prefer laptop B . Imagine that instead of introducing laptop C , a laptop D is introduced into the choice-set with the attributes shown in Choice-set 3:

73 See Joel Huber, John W. Payne and Christopher Puto, “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis”, 9(1) *Journal of Consumer Research* (1982), pp. 90 *et seq*; Itamar Simonson and Amos Tversky, “Choice in context: Tradeoff contrast and extremeness aversion”, 29 *Journal of Marketing Research* (1992), pp. 281 *et seq*.

Choice-set 3

	A	B	D
Price	\$400	\$300	\$350
Storage	300 GB	200 GB	150 GB

Faced with this set of choices the asymmetric dominance effect is likely to influence your preferences such that while you will disregard *D*, you will now find *B* more attractive than *A*.

So is the asymmetric dominance effect to be considered a nudge? The effect is not one that may be reconciled with the rational preferences of Econs.⁷⁴ That is, the effect of adding an, in principle, irrelevant choice option does not affect Econs. Still, it does affect experimental subjects and hence Humans. According to the principle from behavioural economics, then, it counts as a nudge.

But does our revised definition of nudge align with this conclusion? Obviously the addition of an irrelevant choice option does not rule out or forbid any options from the original choice-set. Hence (1) is not violated. Nor does it significantly change the incentives and hence (2) is respected as well. Instead the influence is made possible because of cognitive bias in individual decision-making. This means that condition (3) is satisfied. Finally, the addition of the irrelevant choice option is an intervention that works by making use of those biases. That is, it is not a case of merely de-biasing choice and thus condition (4) is satisfied.

Accordingly, relative to our revised definition, adding choices in this way is a case of a nudge – not only because it is stipulated so, but because it is consistent with the underlying principle saying that a nudge is any factor that significantly alters or influence the behaviour of Humans, even though it would be ignored by Econs. However, before revising the

definition so as to allow for the addition of rationally irrelevant choice options and bar the addition of relevant ones, we will address an interesting issue.

c. Nudging by Removal of Irrelevant Alternatives

If one may nudge by *adding* an irrelevant choice, then it seems reasonable that the same may be the case by *removing* such an irrelevant choice. For instance, going back to the asymmetric dominance effect: if three laptops are available in the status quo as in choice-set 2 and 3, and we then remove the choice-option *C* and *D*, respectively, are these interventions then instances of nudges?

To answer this question we should, according to the definition, first ask whether the removal of a rationally irrelevant choice option like this significantly alters the behaviour of Humans, even though Econs would ignore it. To determine this we may look at the examples above in reverse. Of course, since some of the experiments on the asymmetric dominance effect are *between-subjects* design, the effect of removing the rational irrelevant options is the reverse of that observed when adding it. But even for within-subjects designs the effect occurs.⁷⁵ We then ask whether these behaviour-changes would be consistent with the standard rationality axioms from which the behaviour of Econs is derived – the answer, of course, is no.

To be precise, according to the standard axiom of decision theory referred to as the *independence axiom*,⁷⁶ rational choices should be independent of irrelevant alternatives. This axiom is sometimes also referred to as *Sen's property* stating that:⁷⁷

if an alternative *x* chosen from a set *T*, and *x* is also an element of a subset *S* of *T*, then *x* must be chosen from *S*.

That is, eliminating some of the unchosen alternatives should not affect the selection of *x* as the best option. In our case of the asymmetric dominance effect the independence axiom implies that Econs should be unaffected by the removal of *C* and *D*. Yet, as the evidence for the asymmetric dominance effect and other decoy effects shows, Humans *are* influenced by this removal. Hence, in such cases, removal of irrelevant choice options would qualify as nudges according to the principle from behavioural economics.

It is in light of this conclusion that the definition of a nudge above needs to be further refined so that

74 Huber, Payne and Puto, "Adding Asymmetrically Dominated Alternatives", *supra* note 73; Simonson and Tversky, "Choice in context", *supra* note 73.

75 See e.g. Huber, Payne and Puto, "Adding Asymmetrically Dominated Alternatives", *supra* note 73.

76 Peterson, *An Introduction to Decision Theory*, *supra* note 57, at p. 99.

77 Amartya Sen, *Collective Choice and Social Welfare* (San Francisco: Holden Day, Inc., 1970), at p. 17.

besides barring only the addition of relevant options, it also allows for the removal of rationally irrelevant ones. While the removal of irrelevant choices as a nudge has no bearing on the other conditions, it calls for revising condition (1) by adding the qualification of *relevancy*, such that:

A nudge is a function of the choice architecture that alters people's behaviour in a predictable way that works independently of

(1) forbidding or adding any **rationally relevant choice** options or

(2) changing their incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth.

(3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

4. Information

Earlier it was mentioned that only three basic elements influence the behaviour of an Econ. Choice options, preferences over the outcomes associated with these, and information or beliefs. So far we have revised the definition relative to choice options and preferences. But what about information? Just like it was the case for the addition of choice options, the original definition of a nudge provided by Thaler and Sunstein says nothing about this. So do information and various forms of communication count as nudges?

a. The Provision of Factual Information

Beginning with the mere *provision of factual information* we may observe that in his recent book *Why Nudge?* Sunstein is quite straightforward on this point:

"Provision of information is certainly a nudge, but it may or may not qualify as paternalistic..."⁷⁸

But is Sunstein right in this? Obviously in the sense targeted here, the provision of factual information is usually intended to influence behaviour without forbidding or adding any relevant choice options or significantly changing incentives. Hence, the provision

of factual information would seem to qualify as a nudge given Thaler and Sunstein's original definition taken at face value. In addition, if offered as well intended advice it should definitely also count as both libertarian and paternalistic.

However, if we look to the principle from behavioural economics saying, "a nudge is any factor that significantly alters the behaviour of Humans, even though it would be ignored by Econs" it becomes clear that, despite Sunstein's assertion to the contrary, the provision of factual information should usually not qualify as a nudge. This becomes even clearer if we look at the two final conditions of the definition that captures this principle:

(3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

For one, the provision of factual information would not necessarily be called for because of cognitive boundaries and biases. It would merely be motivated by a person's limited information. Second, under normal circumstances it would be intended to work in a way that would seek to influence Econs as well as Humans. So while it is possible that the provision of factual information may qualify as a case of libertarian paternalism, it does not qualify as a nudge.

Perhaps a reason why nudges and the provision of information may sometimes be confused is because information may be provided in ways that are not intended as the provision of pure factual information that one did not have access to beforehand. Take *reminding* someone of something, say, that I have an appointment with my dentist as in studies like that of Altman and Traxler⁷⁹. Here, conditions (1) – (3) would be satisfied. It would be an attempt to influence my behaviour in a predictable way without forbidding or adding any rationally relevant choice options or changing my incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth. The reminder would be called for because of my *limited memory*, which in turn may be conceived of as the result of cognitive boundaries given Econs are assumed to have perfect memory –

⁷⁸ Sunstein, *Why Nudge?*, *supra* note 6, at p. 55

⁷⁹ Steffen Altmann and Christian Traxler, "Nudges at the Dentist", (IZA Discussion Papers 6699, Institute for the Study of Labor, 2012).

hence an Econ would never need a reminder. But would condition (4) be satisfied? That is, would the reminder be an attempt at influencing my behaviour in a predictable way intended to work by making use of cognitive boundaries, biases, routines or habits as an integral part of that attempt? That is, am I not just being reminded in a way that could be described as ‘de-biasing’? Following Sunstein reminding does satisfy condition (4) since it makes use of *salience*.⁸⁰ That is, it is not the information as such that causes the behavioural effect or function, but that of making the information salient as a means of memory retrieval. Thus, as with the structuring of incentives, the intervention consists of two components - an informational component and a behavioural function in terms of salience. In fact, this means that the whole so-called *notice paradigm* as described by Ryan Calo qualifies as nudging.⁸¹ This includes the infamous “fly in the urinal” which also satisfies condition 4 by making use of salience; although it is called for due to another cognitive boundary *viz.* limited attention, rather than limited memory. In general, such notice-tactics are part of the nudge approach, although they may often seem like mere provision of information. Still, the general conclusion holds: the provision of factual information in and by itself does not count as a nudge.

b. Arguments and Fallacies

Another tactic that is likely to give rise to similar confusion is that of *rational persuasion* or argument. Does this qualify as a nudge? Well, the answer is evident given the term “rational”. For rational persuasion and logical argumentation (1) and (2) will be satisfied as well as perhaps (3); though (3) will not be satisfied because of ‘limited memory’, but perhaps rather because of ‘limited processing capacity’ in connecting the facts. However, condition (4) will not be satisfied: rational persuasion or argument is not intended to work by making use of cognitive bias as part of the attempt at influencing people’s behaviour in a predictable way. In so far as rational persuasion works – not by providing new premises by means of the provision of factual information, but by connecting these premises into a conclusion – it should at most be considered as a de-biasing strategy.

However, the reason to pick up the subject of arguments here is that not all persuasion is rational. One could make use of *logical fallacies* in order to persuade. So would that qualify as a nudge? Obviously, the use of such fallacies would not convince a perfectly rational agent, that is an Econ. Yet, as is well documented, fallacies are often used to convince and mislead Humans. Hence the use of fallacies to persuade ultimately qualifies as nudging when considering the principle from behavioural economics.

The same conclusion is also reached by applying the revised definition. The use of logical fallacies satisfies conditions (1) and (2) as the intervention would be intended to influence behaviour, broadly conceived of, without forbidding or adding any rationally relevant choice options or changing incentives. Likewise conditions (3) and (4) are met since persuading someone by means of a logical fallacy would only be possible due to cognitive boundaries and biases and would work by making use of those biases and boundaries as an integral part of the attempt to persuade. Hence, the use of logical fallacies qualifies as a nudge on this definition as well.

Given the conclusions reached in this discussion of information and fallacies, we may now revise the definition of nudge as follows:

A nudge is a function of the choice architecture that alters people’s behavior in a predictable way that works independently of

- (1) forbidding or adding any rationally relevant choice options,
- (2) changing their incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, **or**
- (5) the provision of factual information or rational argumentation.
- (3) Nudges are called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and (4) work by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

5. Conclusion: Minimal Definition of Nudge

To reconcile the definition of nudge provided by Thaler and Sunstein with the principle from behavioural economics we have revised most elements of

80 Sunstein, *Why Nudge?*, *supra* note 6, at p. 39-44.

81 Calo, “Code, Nudge, or Notice”, *supra* note 7.

the definition. However, at this point one may have noticed that the negative conditions concerning options, preferences, and information are derivatives from this principle, rather than *vice versa*. To emphasise this as well as pre-empt the possibility that even further negative conditions may be identified, I suggest the following re-arrangement of what I put forward as the *minimal definition* of a nudge:

Minimal Definition

A nudge is a function of the choice architecture that alters people's behaviour in a predictable way (1) that is called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of the choice architecture.

Thus a nudge amongst other things works independently of:

- (i) forbidding or adding any rational relevant choice options,
- (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or
- (iii) the provision of factual information or rational argumentation.

V. Nudge and Libertarian Paternalism

Despite having reached a minimal definition of nudge, it is not yet possible to determine the exact relationship between nudge and libertarian paternalism. Obviously a clear overlap exists between these two concepts as it follows from the definition of a nudge that nudges work independently of (i) - (iii). However, the question remains whether any particular motivation, if any motivation at all, is to be tied to nudges and how to deal with this conceptually in order to achieve the wanted clarity and consistency. The purpose of this section is to examine that question.

1. Intentionality (A Nudge, or "to Nudge")

A first answer to the question is found in the discussion about whether the concept of nudge should be treated as a *noun* or a *verb*.⁸² Initially, this may seem a ridiculous point to make, but it turns out to have

some important bearings when one turns to theoretical issues as well as discussions of the ethics of nudge.

In Thaler and Sunstein's original definition, the concept of 'nudge' is treated as a noun referring to "any *aspect* of the choice architecture" (my emphasis). Thaler and Sunstein's formulation of the principle from behavioural economics takes the same stance since, there, a nudge is referred to as "any *factor* that significantly alters the behavior of Humans, even though it would be ignored by Econs" (my emphasis).

However, as in the Monty Python sketch, a "nudge" usually serves as a verb designating an action. Hence, "to nudge" is an action intended by a subject. This also makes it possible to say, "*nudging* the ball" just as you would say, "hitting the ball". Likewise, as Thaler and Sunstein say, libertarian paternalists nudge, meaning that a libertarian paternalist may be nudging. Of course, using the term nudge in this way does not exclude derivative uses implying a lack of intentionality, such as "he *hit* the wall" in the more unfortunate sense. But notice, using the term this way would be *derivative* and not primary (like when we say that "she unintentionally poisoned him", we do not mean that she actually *poisoned* him). Also, intentionality does not imply that when someone "nudges" he necessarily does so with this concept in mind whether in *sensu composito* or in *sensu diviso*. You don't need the concept of 'bragging' to actually do this, nor do you need to intend to 'brag' in individual instances in order to do it.

From the title of their paper "Debate: To Nudge or Not to Nudge"⁸³ Hausman and Welch actually seem to take the "verb" approach to the question. However, on a closer read, it turns out that they oscillate between treating "nudge" as a verb and as a noun. In general they treat nudges as "factors".⁸⁴ Yet, in their summary they refer to nudges as "ways of influencing choice". This formulation may in turn be understood as a noun referring to systematic relations (bias, boundaries, etc.) in an objective world (as when saying "ways in which gravity affects the earth"), or as a verb describing the use of systematic competences

82 See Hansen and Jespersen, "Nudge and the Manipulation of Choice", *supra* note 25.

83 Hausman and Welch, "Debate: To Nudge or Not to Nudge", *supra* note 22.

84 See e.g. *Ibid*, at p. 126.

involving some kind of intentionality. Either way, as they point out by the end of their paper there remains an important difference between choices that are intentionally influenced by a third party, and choices that are not.⁸⁵

It is for reasons resulting from this last point that Hansen and Jespersen argue for consistently treating the concept of nudge as a verb:

... there seems to be a clear and important distinction to be made between a given context that accidentally influences behaviour in a predictable way, and someone – a choice architect – intentionally trying to alter behaviour by fiddling with such contexts.⁸⁶

The reason why Hansen and Jespersen make this point is that in matters of normative justification one simply cannot dispense with the issue of intentionality and, by extension, agency. Intentionality is a conceptual precondition of normative evaluation. Ignoring it would render the notion of responsibility superfluous. Unfortunately, as Hansen and Jespersen also point out, this is exactly what is often indicated when treating the concept of “nudge” as a noun, rather than as a verb. Such a conceptual move blurs a crucial distinction at the heart of normative justification as to the notion of responsibility. Hence the seemingly ridiculous question of whether to treat the concept of “nudge” as a noun or a verb makes a crucial difference because it introduces intentionality and thus the normative dimension of responsibility into the debate about the nudge approach to behaviour change.

For this reason Hansen and Jespersen argue for stipulating the definition of nudge as being one of a verb in order to ensure that this important distinction is not lost.

Thus, we suggest that a nudge henceforth is best understood as the intentional attempt at influencing choice, while it is accepted that the settings of any given decision-making context may accidentally influence choice and behaviour in predictable ways as well. ... The notion of “nudge” then, should only apply when someone intentionally tries to in-

fluence our behaviour without the use of regulation or fiddling around with incentives.⁸⁷

So returning to the revised definition we may now follow Hansen and Jespersen 2013 by integrating a more precise and viable condition substituting the passive aspect of “choice architecture” with condition (I) stipulating that a nudge is a function of any (*intentional*) attempt at influencing people’s behaviour. While modifying this part of the definition we may also note that it may be beneficial to make explicit what is implicitly contained in the concept of behaviour - namely judgment, choice as well as overt behaviour - in order to avoid confusions on that point.

A nudge is a function of (I) any **attempt at influencing people’s judgment, choice or** behaviour in a predictable way (1) that is called for because of cognitive boundaries, biases, routines, and habits in individual and social decision-making, and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of **such attempts**.

Thus a nudge amongst other things works independently of:

- (i) forbidding or adding any rationally relevant choice options,
- (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or
- (iii) the provision of factual information or rational argumentation.

As we will see below, this at the same time clarifies why the terms nudge and bias are not logically equivalent and this hence precludes a logical substitution that otherwise would make for a conceptual oddity.

2. Who’s Calling? Or, Why Nudge?

So far we have determined that a nudge is an intentional attempt at influencing people’s behaviours in a predictable way under certain constraints. However, this raises the next question of whether a special motivation for forming this intention is required as well in order for such an act to qualify as a nudge. For instance, is a libertarian paternalistic motive required, or could other motives do as well? The addition by means of condition (1) in the revised definition adopted might indicate so by stating that

85 Ibid, at p. 133.

86 Hansen and Jespersen, “Nudge and the Manipulation of Choice”, *supra* note 25, at p. 10.

87 Ibid.

“nudges are called for” – but what position on this provides consistency and clarity?

Hausman and Welch’s formulation “nudges are called for” definitely seems to invite an interpretation in terms of the *motivation* behind the nudge. That is, it seems to refer to someone’s reason for intervening – the motive behind the (I) attempt at nudging. For this reason we may revise condition (1) so that it says:

- (1) that is **motivated** because of cognitive boundaries, biases, routines, and habits in individual and social decision-making

But what kind of motive, if any, is required by the definition? Here, Hausman & Welch’s attempt at a definition says that nudges are called for “because of flaws in individual decision-making”.⁸⁸ This means that condition (1) may be read as identifying a *pre-condition* for when a nudge is called for: *given* cognitive boundaries, biases (recall that we dropped the concept of ‘flaws’), routines, and habits in individual and social decision-making, nudges are called for. But why would that be? One possible and very likely reason would be that since boundaries, bias, routines, and habits may prevent one from acting according to the standards of rationality *then* nudges “are called for” because these biases pose barriers for people to perform (act, judge, evaluate) rationally. If we follow this interpretation we might thus consider revising condition (1) such that we get:

- (1) that is motivated because of cognitive boundaries, biases, routines, and habits in individual and social decision-making **posing barriers for people to perform rationally in their own declared self-interest.**

Now this addition explicates the notion of cognitive boundaries and biases, as well as certain routines, and habits. But why does the occurrence of cognitive boundaries, biases, routines, and habits result in a reason to nudge? That is, who or what is calling for the need to nudge? As it turns out, a lot hangs on the interpretation of this subtlety only implicit even in Hausman and Welch’s attempt at explicating the definition provided by Thaler and Sunstein. I find that two possible interpretations are readily available.

a. Nudging as Libertarian Paternalism

First, we may interpret the definition as requiring that the motive be *paternalistic*. That is, because of

cognitive boundaries and biases in individual and social decision-making, the nudged agent is prevented from acting rationally according to his own interest as judged by the one nudging or as judged by himself – and this may be taken to *call* for a paternalistic intervention, which as a matter of definition is done in the interest of the people nudged (as judged by the one nudging or as judged by himself). This is the position taken by Sunstein in his recent *Why Nudge?*⁸⁹ If this interpretation is chosen, we have that:

A nudge is a function of (I) any attempt at influencing people’s judgment, choice or behaviour in a predictable way **according to their own self-declared interests** (1) that is motivated because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally **in their own self-declared interests**, and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts. Thus a nudge amongst other things works independently of:

- (i) forbidding or adding any rationally relevant choice options,
- (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or
- (iii) the provision of factual information or rational argumentation.

If choosing this interpretation of the motive in (1) we essentially marry the concept of nudge to that of libertarian paternalism. Hence we may refer to this definition as *the LP-definition*. Under this definition nudges are a subset of libertarian paternalism. See Figure 2.

b. Nudging in the Technical Sense

Alternatively we may interpret the definition as not making any specific requirements as to the motive, other than that the motive is dependent on or made possible by the fact that cognitive boundaries, biases, routines, and habits in individual and social decision-making pose barriers for people to perform ra-

88 Hausman and Welch, “Debate: To Nudge or Not to Nudge”, *supra* note 22, at p. 126.

89 Sunstein, *Why Nudge?*, *supra* note 6.

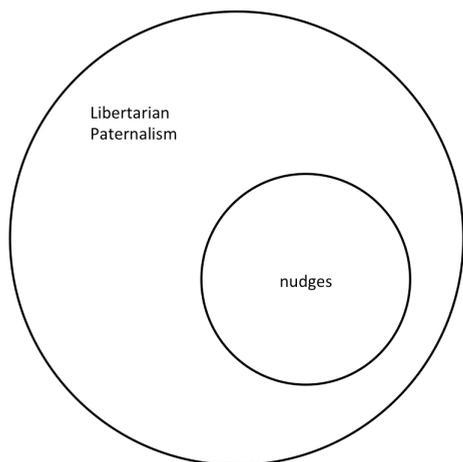


Figure 2

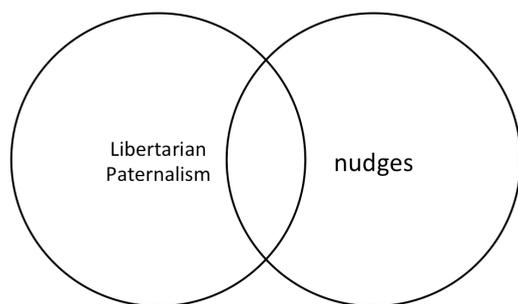


Figure 3

tionally. If we choose this strictly *technical* interpretation we get a definition saying that:

A nudge is a function of (I) any attempt at influencing people's judgment, choice, or behaviour in a predictable way that is (1) made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally in **their own self-declared interests**, and which (2) works by making use of those biases, routines, and habits as integral parts of such attempts.

Thus nudges amongst other things work independently of:

- (i) forbidding or adding any rationally relevant choice options,
- (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or
- (iii) the provision of factual information or rational argumentation.

Obviously, "made possible" is weaker than "called for" and thus, different from the LP-definition, the technical definition is not married to the notion of libertarian paternalism. Hence we refer to it as the *technical*, or the *neutral* definition. Under this definition the set of nudges overlaps with that of libertarian paternalism, but is not contained in the latter. See Figure 3.

Now, to decide which definition to prefer, one could turn exegetic. However, in my opinion, re-reading what has been written by Thaler and Sunstein about nudges to search for clues in one or the other direction would not lead to any decisive conclusions. Whether they favour the LP-definition or the technical definition does not decide the question of which stipulative definition we should opt for. However, as I have argued elsewhere there is one set of exegetic readings that lend evidence to the technical definition as the best *stipulative* definition of the concept of nudge given the way that this concept has come to be adopted in discussions surrounding the nudge approach to behaviour change.⁹⁰

c. "Nudge for Good"

First, the technical definition is consistent with certain central comments made by Thaler and Sunstein.

In general Thaler signs copies of *Nudge* with the slogan "nudge for good". This precautionary call clearly indicates that the notion of 'nudge' is not necessarily married to that of 'libertarian paternalism', but may instead be cast in terms of the technical definition – because if this weren't the case, Thaler's phrase would be nothing more than a tautology. Something similar goes for the title of Sunstein's recent book *Why Nudge?* After all, if the LP-motive were part of the definition of a nudge, then one would not have to answer the question of 'why nudge?' – one would only have to define the term. In addition one may observe that when Sunstein says that the "Provision of information is certainly a nudge, but it may or may not qualify as paternalistic..."⁹¹ this seems to reveal

90 Pelle G. Hansen, "Nudge for good", Policy Options, 3 Jun 2013, pp. 22-23 et sqq.

91 Sunstein, *Why Nudge?*, supra note 6, at p. 55.

that nudges that fall outside the scope of libertarian paternalism exist and hence the set of nudges cannot be a subset of libertarian paternalistic measures. These considerations support the adoption of the technical definition.

Second, the technical definition is consistent with points about nudges and marketing consistently made by commentaries. If we opt for the technical definition it makes sense to refer to the many marketing tricks used to fool us into buying things we don't need as nudges as well. That is, the technical definition recognises that it is possible to nudge "for bad" or "for profit" as well as "for good". This observation is both in line with many comments made in the general literature and mentioned here in the introduction as well as many formulations in the work of Thaler and Sunstein themselves.

Thirdly, when the nudge approach to behaviour change is regarded as a sub-branch of libertarian paternalism or as synonymous with this, there is a danger that anyone in favour of the former must also be in favour of the latter. That is, if one sees nudges as valuable measures for creating behaviour change, one is often taken as favouring the political ideology of libertarian paternalism - an ideology that by some is said to hold that policy makers should avoid regulations that limit choice (bans, caps, etc.) but can use behavioural science to direct people towards better choices. However, this is not necessarily so on the technical definition. Here one may adopt nudges as valuable measures to behaviour change without adopting a particular political ideology.

There are probably other reasons for sticking with the technical definition of nudge. However, I will accept the three mentioned above as sufficient as they lend evidence to the technical definition as the best *stipulative* definition of the concept of nudge, refining what has come to be standard usage in the reception of *Nudge*. That is, while one may choose either one as one's favourite definition, it seems that there

is a conceptual need for adopting the technical one since it is better at providing clarity when discussing applications of behavioural economics and other behavioural sciences with the aim of creating behaviour change. For this reason I conclude that the most suitable definition of nudge is as follows:

A nudge is a function of (I) any attempt at influencing people's judgment, choice or behaviour in a predictable way, that is (1) made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally in their own self-declared interests, and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts.

Thus a nudge amongst other things works independently of:

- (i) forbidding or adding any rationally relevant choice options,
- (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or
- (iii) the provision of factual information and rational argumentation.

In so far as a nudge serves the self-declared interest of those being nudged, it may further be referred to as *libertarian paternalism* since the general definition of nudge implies that people's judgment, choice, or behaviour is influenced in ways that work independently of (i) forbidding or adding any rationally relevant choice options, (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economic incentives and so forth. In addition, this definition also implies that libertarian paternalism goes beyond nudging since it follows from it that nudges work independently of (iii) the provision of factual information and rational argumentation, which fall squarely within libertarian paternalism.