



# From mechanism to virtue: Evaluating Nudge theory

**Mark Kusters**

Australian Government's Office of Best Practice Regulation, Australia

**Jeroen Van der Heijden**

Australian National University, Australia and University of Amsterdam, The Netherlands

## Abstract

Ever since Thaler and Sunstein published their influential book *Nudge*, the book and the theory it presents have received great praise and opposition. Nudge theory, and more particularly, nudging may be considered an additional strategy providing some novel instruments to the already rich governance toolbox. But what is its value? The current debates on Nudge theory are often highly normative or ideologically driven and pay limited attention to more practical aspects of the theory. Is nudging evaluable as a theory and a practice, and if so how? Is there solid evidence available of nudge success over other governance interventions? What is to be considered a nudge success? What data and evaluative techniques may assist in evaluating nudging beyond individual cases? The current article seeks to explore these questions.

## Keywords

governance, governance evaluation, governance instruments, Nudge theory, policymaking

## Introduction

It is not often that a book on governance interventions becomes a bestseller, but Richard Thaler and Cass Sunstein's *Nudge: Improving Decisions about Health, Wealth and Happiness* ('Nudge') has. Published in 2008,<sup>1</sup> the book outlines the seemingly immense potential for more innovative and less coercive government interventions to shape people's behaviour. It was considered to be 'Best Book of the Year, 2008' by both *The Economist* and *The Financial Times*. This is an extraordinary achievement compared with all the books on governance interventions published each year that remain in their publishers' warehouses and get read by, at best, a handful of in-group specialist scholars.

---

## Corresponding author:

Jeroen van der Heijden, Australian National University, Regulatory Institutions Network (RegNet), Canberra, ACT 0200 Australia.

Email: [j.j.vanderheijden@anu.edu.au](mailto:j.j.vanderheijden@anu.edu.au)

The success story continues. Not only is *Nudge* well read, it is also well applied. As former Administrator of the White House Office of Information and Regulatory Affairs (OIRA) in the Obama administration, Sunstein has had much opportunity to spread his ideas and apply them to real-world policy matters.<sup>2</sup> On the other side of the Atlantic, in 2010, the UK Government installed the Behavioural Insight Team (BIT), or ‘Nudge Unit’, to apply insights from academic research in behavioural economics and psychology to public policy and services (Thaler acts in an advisory capacity in the BIT). In short, the BIT builds on Nudge theory to improve public policy and services (BIT, 2013). Nudge theory has taken such an enormous flight in the UK that some critical scholars already talk of a ‘nudgeocracy’ (Whitehead et al., 2012: 303) – although these claims seem a little far reaching, as nudging does not appear to be a dominant governance strategy in the UK (Dorey, 2014; Richards et al., 2014). On mainland Europe, Nudge theory has gained attention in France, Germany and from the European Commission, with the latter having installed a Behavioural Studies for European Policies Program in 2010. Interestingly, the UK BIT in its turn has inspired the Obama administration to install its own Behavioral Insight Team, which is already nicknamed the ‘Nudge Squad’ by the US media (Fox News, 2013).

In reading the background information on this UK Unit, EU Program and US Squad, one comparable issue stands out: it is considered that Nudge theorizing ‘allows policy-makers to better understand and influence people’s behaviour’ (European Commission, 2013), which will ‘help [to] design public policies that work better, cost less, and help people to achieve their goals’.<sup>3</sup> These oft-repeated expectations on nudging by policymakers and administrators are especially striking because the evidence base of its success as documented in the academic and grey literature is, to say the least, disputed, as we will further discuss in this article. Nudge theory appears to have started from empirical observations of a series of experiments with governance instruments (a mechanism), but has quickly achieved the normative status of a governance wonder nostrum (a virtue) able to solve problems arising from earlier direct and coercive governance interventions. Such a move from a governing mechanism to a governing virtue is not unique to Nudge theory. A similar trend can be uncovered for, for instance, the new public management movement (Pollit, 1995), which was popularized by the immensely influential book *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* by David Osborne and Ted Gaebler (1992).

There is, of course, nothing wrong with policymakers believing in the virtues of a new type of governance intervention, but we argue for some care in considering nudging to be a virtue only. That is not to say that we rank ourselves amongst those who are very critical of Nudge theory and consider it to, potentially, do more harm than good (e.g. Bradbury et al., 2013; Goodwin, 2012; Pinch, 2010; Selinger and Whyte, 2011). We argue that Nudge theory lacks a sufficient evidence base for either of these somewhat polarized points of view. We further argue that Nudge theory lacks a solid evaluative framework, which makes it difficult to assess whether nudging as a governance intervention outperforms other governance interventions, under what circumstances and why.

Preparatory to future evaluation of real-world scenarios of nudging, this article therefore seeks to answer four related questions: Whether and how is nudging evaluable as a theory and a practice? What may be considered nudging success? To what extent is solid evidence available of nudge success? What data and evaluative techniques may assist in evaluating nudging beyond individual cases? In order to answer these questions, we first seek to better understand Nudge theory, critiques of the theory, and first insights into the performance of nudging in real-world settings.

## What exactly is a nudge?

The central premise behind Thaler and Sunstein's nudges is that individuals are not 'Econs'. In other words, people frequently behave in a way that economic theory finds difficult to predict (Thaler and Sunstein, 2009: 7). It is, of course, not a novel insight that people do not act in their own best interest from an economics point of view. In 1973, Herbert Simon was awarded a Nobel Prize for his classic work on bounded rationality (Simon, 1945), which argues that people are unable to make economically optimal decisions because they lack the capacity to store the voluminous information needed for such decisions, as well as the cognitive ability to process that information. More recently, influential behavioural scientists such as Daniel Kahneman (2011), Roberto Cialdini (2009) and Dan Ariely (2008) have evidenced what for a long time has been common knowledge: people do things that are not in their own best interest even when they are aware of this (bounded willpower), or that people do so because they consider benevolent behaviour as more fair than selfish behaviour (bounded self-interest).

Despite this long history of pointing out that 'Homo Economicus is a fiction' (Leonard, 2008: 356–7) law, and specifically the economic analysis of legal rules, has been hampered by its focus on the substance of the rules (i.e. how they are written) and their function (i.e. how they should operate) rather than being concerned with how people respond to them (Bardach and Kagan, 1982; Jolls et al., 1998; Sparrow, 2000; Supiot, 2007). In an analysis that predates *Nudge*, Thaler and Sunstein, along with Christine Jolls, suggest that while behavioural insights challenge the 'simplicity and parsimony' of conventional economic analysis of legal rules, they can provide a more accurate prediction of the likely behavioural response to legal rules (Jolls et al., 1998: 1487–8).

In *Nudge*, the focus shifts from employing behavioural insights as a prognosticator of human behavioural response to being key in the practice of shaping that behaviour. This application of behavioural insights is what Thaler and Sunstein refer to as changing the 'choice architecture' of those governed. To them, the answer to the question 'why nudge?' is that the application of behavioural insight should result in more effective governing and, in doing so, make people happier. For government, the answer is similar in some respects, in that effective governance, policies and regulations are obviously preferred over ineffective ones. *Nudge* also represents a welcome respite from the inherent tensions of a deregulatory agenda, which is often criticized for not having achieved its desired outcomes (see John and Richardson, 2012). This all has created a call for innovative governance interventions (Van der Heijden, 2013), and we consider nudging one such intervention. Further, in the current climate of fiscal restraint, *Nudge's* potential to offer 'low cost' solutions to some of the more complicated governance problems has been a strong factor in winning the support of the US and UK governments, respectively (BBC Radio 4, 2013).

## *Nudge: A somewhat flexible definition*

*Nudge* presents a series of truisms and well-selected research examples to illustrate that individuals are not 'Econs', and to support the central claim of the book that changes in 'choice architecture' may help them to improve their decisions about health, wealth and happiness. But what exactly is a nudge? Thaler and Sunstein provide us with a somewhat flexible definition in the introduction to the book:

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 cheap and easy to avoid. Nudges are not mandates. Putting fruit at eye level [hoping that people then choose fruit over unhealthy alternatives] counts as a nudge. Banning junk food does not. (Thaler and Sunstein, 2009: 6)

Throughout the remainder of the book, Thaler and Sunstein (2009) describe a nudge as the application of insights from behavioural economics in government policy as a substitute for more conventional coercive interventions such as command and control regulation. This makes their work sit in the broader governance literature that is interested in new and innovative instruments that seek to steer ('govern') human behaviour towards desired ends (for reviews of the literature, see Bell and Hindmoor, 2009; Chhotray and Stoker, 2010). The examples selected by Thaler and Sunstein do, however, cause more confusion than clarity about what a nudge *exactly* is.

Looking at the examples discussed by Thaler and Sunstein, it becomes clear that nudges can be implemented by a wide range of actors, come in a wide range of forms, and seek to achieve a wide range of outcomes. For one, it seems that anyone can nudge – government can nudge (e.g. to lower rates of teenage pregnancy, governments may pay teenage girls one dollar each day they are not pregnant), business can nudge (e.g. to improve health outcomes and reduce pay-outs, medical insurers may offer financial incentives for members who undertake prescribed activities to improve their health) and individuals can nudge themselves (e.g. in order to quit smoking, a person can set aside a given amount of money for six months and only gets the money returned if a test for nicotine is passed; if not, the money is donated to charity). Also, nudges seek to change behaviour through a wide range of interventions. They can seek to change behaviour through a financial incentive (as per the above examples of medical insurance, or teenage pregnancy), through providing relevant information (e.g. to improve energy efficiency, air conditioners come with a bright red light to signal that the filter needs to be changed) or even by actively blocking an inappropriate choice (e.g. to prevent sending an email that the sender will later regret, computers filter outgoing messages for strong language, which if detected blocks the message for 24 hours). Finally, nudges may seek to achieve substantially different outcomes. They can provoke a single response (i.e. change the air conditioner filter, rewrite an email), prompt a more long-lasting behavioural change (quit smoking) or both (i.e. sign up for gym membership to improve health).

### ***Betting on two horses?***

What strikes us is that the common, distinguishing feature of these nudges is not clear. There does not seem to be an obvious common denominator that truly defines nudges as a different approach to steering individuals' behaviour than other governance interventions – may these be implemented by governments, non-governments or a collaboration of both (for examples, see Chhotray and Stoker, 2010). How exactly do nudges based on financial incentives differ from traditional governmental interventions such as subsidies, grants or taxes? How exactly do nudges based on blocking an inappropriate choice differ from traditional restrictive regulation? How exactly are nudges based on information supply different from a range of informational instruments that have been trialled with by governments for decades now?

This heterogeneity of Thaler and Sunstein's definition has resulted in some criticism of an overstretching and fuzziness of nudging as a concept (e.g. Hausman and Welch, 2010; Selinger and Whyte, 2011). In addition, nudging as discussed by Thaler and Sunstein can be one of two things. First, it can be a governance intervention that seeks to assist individuals to make choices that are in their own best interest (i.e. saving more for retirement, choosing the right insurance plan) – referred to as 'type 1 nudges' in what follows. Second, it can be a governance intervention that seeks to steer

individuals' behaviour to achieve a desired collective end (e.g. reducing crime, encouraging environmentally-friendly practices) – referred to as 'type 2 nudges'. Of course, there are synergies between the two (i.e. individual benefits such as quitting smoking may reduce future government outlays on interventions, such as the cost of healthcare), yet distinguishing among them will be of essence in evaluating nudging. The two interventions, after all, seek to achieve significantly different goals.

To say the least, all this vagueness will not help policymakers (and academics alike) to gain a better understanding of where, when and why nudging may be expected to result in desired outcomes. Nor is it likely that it will help to better understand whether nudging results in more optimal outcomes than other governance interventions. In other words, nudging as defined and illustrated by Thaler and Sunstein is difficult, if not impossible, to evaluate because of their seemingly all-inclusive understanding of the governance intervention they seek to capture. Scholars seeking to evaluate nudges are, we feel, forced to redefine the exact meaning of nudging in their own studies.

### **Nudge critique, or are these unevaluated fears?**

Nudge theory has not been without its critics (e.g. Bradbury et al., 2013; Goodwin, 2012; Pinch, 2010; Selinger and Whyte, 2011; White, 2013). Some have warned of its capacity to 'sow the seeds of an illiberal system of control' and that its effectiveness relies on factors which are difficult to predict (Baldwin et al., 2011: 124–5). In 2011, the UK Science and Technology Committee concluded that regulatory experiments, such as those being trialled by the BIT, were unlikely to be effective in isolation and did not appear to be supported by appropriate evidence (Science and Technology Committee, 2011). In a follow-up report on the BIT in 2014, the Committee states to be 'impressed by their work in pioneering the use of randomised controlled trials to determine the effectiveness of [behavioural] interventions', but is still concerned 'that robust data [on the effectiveness of such interventions]' is not always available in the public domain to substantiate [claims made by the Unit on the success of such interventions] (Science and Technology Committee, 2014: 1). BIT Director David Halpern observed that the initial resistance from the UK bureaucracy and media to his work in the Unit has abated, but there remain areas within the UK government which do not support it (Halpern, 2012). Further, Sunstein has since resigned from his OIRA position, with some suggesting that his efforts to 'test his theories of human behaviour and economic efficiency in the laboratory of the federal government' have left both business and consumer groups disappointed and done 'untold damage' (Broder, 2012). Although the effectiveness of nudging as a governance intervention is questioned (see also Section 4), Nudge theory is mostly critiqued for its political stance. Most fierce critiques address the legitimacy, accountability and transparency of nudging. But what is the evaluative base of these critiques?

### ***Nudging legitimacy***

Setting defaults and making decisions on behalf of citizens is not without risk. The most common critique is that Nudge theory, and especially its underlying political philosophy of libertarian paternalism, violates the principles of individual freedom. After all, under a nudge regime individuals are no longer able to make their own choice, so goes the arguments of critical scholars, but are nudged in making the 'right' choice by government (Bradbury et al., 2013; Brown, 2012; Goodwin, 2012). In these situations, the choice depends on a subjective assessment (the choice architect's 'conception of the good'), which may be subject to personal bias and motivations

(Baldwin et al., 2011). A counterargument is, of course, that any policy that seeks to change behaviour or a societal desired collective end will have to make choices about the type of behaviour desired, the social end required and the governance intervention that is expected to be most likely to achieve such goals. In other words, any policy and its related governance intervention(s) will have to set out a choice for the citizen – even when developed in close collaboration with the citizens it seeks to govern (Van der Heijden and Ten Heuvelhof, 2012).

Another concern relates to the quantum of choices made available to decision makers (i.e. those nudged) and affected by those in governing positions (i.e. those nudging). In *Nudge*, Thaler and Sunstein rail against the policy design for a Swedish savings programme and a US prescription subsidy plan. Both plans sought to provide individuals more choice than they had under former governance interventions (see further Section 4). Thaler and Sunstein (2009) claim that both interventions were seriously flawed due to what they call the ‘just maximise choice’ mantra (p. 161). Based on these two carefully selected examples, the inference is made that limiting choice is desirable over maximizing choice. This, however, raises questions unanswered in *Nudge*. For example, is there a role for government in restricting the number of options on offer for a particular product (through regulation, or another governance intervention) or should it be self-regulated by market players? If so, how would this restriction be implemented and enforced? What is the right number of choice options? Further, what are the implications of restricting choice on the market (e.g. less competition, reduced innovation) and for consumers (i.e. some consumers may find they are no longer able to find products that meet their specific needs)? While ‘maximizing choice’ may paralyse those subject to nudging as a governance instrument, the alternative of limiting choice appears problematic as well. Also, research has pointed out that choice architecture does not work out similarly for governed actors from different socio-economic backgrounds – that is, less affluent actors appear to make poorer choices than affluent actors (Bradbury et al., 2013). This calls into question whether nudging is an inclusive or exclusive governance intervention.

### *Nudging accountability and transparency*

Besides, the covert nature of nudges calls has raised questions as to whether nudges provide sufficient transparency and accountability (Baldwin et al., 2011). For example, if governments employ Nudge-type instruments to influence choices at a subconscious level, particularly to prompt the governed actors to choose something other than what they would normally have chosen, it may be counter-intuitive to call these actors’ attention to governments’ intent or method. BIT Director Halpern acknowledges this weakness by arguing that ‘tacit acceptance’ is required from the public (Halpern, 2012). Yet, once the public has given permission to be nudged, government has a broad mandate which obviates the need to seek separate permission for each nudge. This requires considerable trust on the part of the citizen that such a mandate will not be misused. Yet again, this holds true for almost any government-led governance intervention.

Thaler and Sunstein (2009) also argue that transparency is essential. The key to this is constraining nudges to governance interventions that government must be able and willing to defend publicly. However, what the government and its constituents consider ‘defensible’ may not accord. Without disclosure this cannot be contested. Without transparency, the accountability of nudging is also suspect. This is not just about the absence of disclosure, but also ‘the possibility of debate, of questions by the forum and answers by the actor, and eventually of judgement of the actor by the forum’ (Bovens, 2010: 951). This absence of such safeguards can reduce incentives for bureaucrats to fix mistakes (thereby entrenching poor policy) and, more importantly, for ‘public officials or public organisations [to] remain on the virtuous path’ (Bovens, 2010: 954).



### *Evaluating the critique: Do the fears hold in the real-world?*

Although we agree with such criticism, our stance is, as already highlighted, that it applies to other governance interventions as well (see Baldwin et al., 2011; Bell and Hindmoor, 2009; Power, 1999). In both highly coercive governance interventions, such as command and control, and low coercive governance interventions, such as government-led voluntary programs (Crocì, 2005; Lyon and Maxwell, 2000), those in a governing position make choices on behalf of those governed. We consider nudging as ‘just another’ governance strategy providing some additional instruments to the governance toolbox. The application of any of these, nudge-based or otherwise, needs careful consideration, evaluation and discussion with the public to ensure legitimate use. Further, to date, nudging has been confined to, at the very best, a limited role. The criticisms in the literature to nudging legitimacy, accountability and transparency do have a point, but the extent to which it is critiqued (e.g. by considering the UK a ‘nudgeocracy’, Whitehead et al., 2012: 303) is out of step with the real-world implementation of Nudge theory. Besides, as we further discuss below, in real-world settings, predominantly type-2 nudges are implemented (i.e. those seeking to achieve a desired collective end), and often only as an *addition* to more traditional governance interventions such as direct regulation, subsidies or taxes. In other words, the fears make for great opinion pieces and thought experiments, but they hardly materialize in real-world settings.

That having been said, the critical discussions on nudging provide for some intriguing evaluation criteria and questions. First is the question of how effective and efficient nudges are in achieving their desired ends. But other questions are of relevance as well. How is nudging legitimized in real-world settings? How transparent and accountable are nudging strategies in real-world settings? The readers of this journal will be very aware of the literature that discusses the complications of using these various criteria in assessing any policy, or governance strategy and instrument. However, it also points out a vantage point to address this challenge (e.g. Armstrong, 2005; Bovens, 1998; Bovens et al., 2008; Hodge, 2000; Lundqvist, 1988; Power, 1999; Verhoest et al., 2007).

### **Evaluating nudging: Success and failure**

An area where one would expect active evaluation of nudging is whether nudges indeed result in their desired outcomes. Indeed, Nudge theory has been the centre of a range of tests, trials and evaluations. Table SD1 (See Supplementary Material on <http://evi.sagepub.com/supplemental>) provides a summary of those we discuss here.

#### *Success stories*

In reviewing the literature on nudging trials and experiments, a range of nudging success stories can be uncovered. To give a mosaic of findings, the example frequently cited is a comparative study of organ donation arrangements across European countries (Johnson and Goldstein, 2003). This study showed that countries where donors are given the choice to ‘opt out’ (i.e. consent to donate is presumed) have significantly higher proportion of donors – ranging from 85.9 percent in Sweden to 99.98 percent in Austria – compared to countries where donors are given the choice to ‘opt in’ – from 4.25 percent in Denmark to 27.5 percent in the Netherlands. A BIT trial has evaluated the use of text messages on the repayment of court fines, including whether different wording affected response rates. The trial found that personalized information (i.e. including the person’s name) resulted in a higher response than a standard message (33% compared to 25%) and higher repayments (by approximately 30%). The BIT’s theory is that personalized information increases the salience of a message, thereby improving the likelihood of the desired response (Gallagher, 2012).

A large trial in the USA found that positioning the signature box at the start of a form (rather than the end) prompted a more honest response (Cabinet Office, 2012). In this case, customers were required to report the number of miles on an insurance form (the more miles reported, the higher the premium). Those required to sign at the start of the form reported travelling 10 percent more miles than those who signed at the end. Another BIT trial found that simplifying letters could significantly increase response rates (35% compared to the generic letter's result of 4%). In this case, letters prompting doctors and dentists to pay outstanding tax liabilities were found to be more effective if the consequences of non-payment were made more explicit (rather than buried in the text), key messages were highlighted and the letter made shorter (Gallagher, 2012).

Further, women were found to be more concerned about developing breast cancer when informed about their comparative risk as distinguished from their absolute risk (Lipkus et al., 2005). Students were found to be more likely to choose vegetables when these were presented attractively (Hanks et al., 2013). Or, giving consumers feedback on their relative energy consumption in the form of a card with a 'smiley' for those who consumed less than average and a 'frownie' for those who consumed more than average was found to result in overall energy reduction (Schultz et al., 2007). A similar trial encouraging people to pay their taxes in the UK found that the basis of the norm was also significant. Specifically, higher responses were received when households' performance was compared against the town average (83%) than if compared against a broader population, such as the average for the postcode area (79%) or nation (72.5%) (Cabinet Office, 2012). Other examples come from the fields of transportation (Gaker et al., 2010) and food waste (Nomura et al., 2011) and report similar positive results. All these examples build on the explication of biases, assuming that people will respond accordingly when biases are triggered.

Also, demonstrated normative values for one behaviour can indirectly influence other behaviours. A trial in the Netherlands found that people were less likely to litter (33% compared to 69%) and less likely to steal (13% compared to 27%) in an environment free of graffiti (Keizer, 2012). Conversely, signage which highlighted that a certain activity was prohibited (in this case, littering) actually increased levels of that activity (61% without signs compared to 70% with signs), suggesting that, with signage, people become more aware that others are acting inappropriately (Keizer, 2012). Also, people were found more inclined to donate money when triggered to think about something religious (Shariff and Norenzayan, 2007). In another trial, people were found more inclined to help a passer-by if they observe someone nearby sweeping a street (Keizer, 2012).

### *Stories of lesser success*

Despite the mosaic of success stories just presented, some studies present less positive results. An interesting example comes from an Israeli study (Gneezy and Rustichini, 2000). Students were asked to participate in volunteering work. Those agreeing were split in three groups. One group was given a financial token reward for volunteering, one group was given a large token reward, and one group was given no reward at all (the control group). It was found that both groups that received a token reward underperformed in their task as compared to the control group, with the group receiving a small token showing a 35 percent underperformance and the group receiving a large token showing an underperformance of 7 percent. Various studies have reported similar results: that rewarding people for desired behaviour that they initially did on a voluntary basis undermined their willingness to continue doing so because they considered the reward insufficient for the effort the desired behaviour takes (Kamenica, 2012; Pink, 2009).

A slightly different example is presented by Thaler and Sunstein (2009). They use two cases to highlight that changing the 'choice architecture' is not always done effectively by governments.



In Sweden, citizens who did not choose their own retirement savings plan were automatically enrolled into a default plan, but the government funded an extensive media campaign urging people to make their own choice. In the USA, rather than selecting a single default on the basis of merit, eligible citizens who did not choose their own prescription subsidy plan were assigned to a randomly selected plan. In both cases, the results were far from optimal. By carefully selecting these two cases, Thaler and Sunstein seek to demonstrate that, even where a default makes sense, governments still want individuals to make their own choice, even at the risk of poor outcomes. According to Thaler and Sunstein, the implication of this is that governments should be less averse to setting defaults and embrace a more paternalistic role in the decision-making process.

Another way of looking at these two examples Thaler and Sunstein use is that different incentive strategies should not be mixed too much. Whilst individually strategies may hold much promise for results, combined they may undo one another's positive outcomes. Nudges may also cause a boomerang effect. For example, households who learn that their energy consumption is lower than the average may start consuming more as their bias of being a (too) good citizen may be triggered. The earlier discussed study that used 'smileys' to reward desired behaviour and 'frownies' to discipline non-desired behaviour found exactly this (Schultz et al., 2007). Even more, the BIT has acknowledged that, on occasion, interventions which they intuitively thought should work were found to be 'ineffective or even harmful' (Haynes et al., 2012: 17). The potential for adverse impacts from nudges was also highlighted by the UK Science and Technology Committee (2011).

### *Shortfalls in current nudging evaluations*

Intriguingly, positive findings on the nudge trials and experiments reviewed are often presented by mentioning the many caveats of the trials and experiments, as well as their limitations. This may partly explain the differences in the reported success stories and the experienced over-all success (or for that matter, lack thereof) of nudging (see Science and Technology Committee, 2014). After all, such warnings may easily be ignored by those in favour for nudging (overlooking warnings that go against one's ideological preferences is not unheard of; Kahneman, 2011; Rose, 2011), whilst they may be overly emphasized by those opposing nudging as a governance instrument (which again is not unheard of; Kahneman, 2011; Rose, 2011). The same holds for using these evaluations' findings. Take, for instance, a 'classic' nudging experiment on students' choice of food in a school cafeteria. In *Nudge*, Thaler and Sunstein (2009) predict that students will choose healthy food over less healthy alternatives if healthy food is presented as the attractive alternative. A recent evaluation of exactly this setting (Hanks et al., 2013) indeed reports considerable success. When vegetables were presented as the attractive alternative over a less healthy alternative, students were 23 percent more likely to choose vegetables than in the control situation (i.e. vegetables not being presented as attractive). At first glance, this is a great success story. Yet, in reading a bit further, it becomes clear that the 23 percent relative increase corresponds with an 8.2 percent absolute increase: in the control situation 35.8 percent of students chose vegetables, while in the test setting 44 percent did – leaving the majority in both groups choosing the unhealthy option. A critic may argue that an 8.2 percent change is marginal at best, whilst proponents may argue that an 8.2 percent increase at hardly any (governance) cost is an impressive success indeed.

One of the main caveats noted is that nudges that are evidenced to work for one group or in one context should by no means be expected to also work for another group or in another context (Hausman and Welch, 2010; Michie and West, 2013; Willis, 2013). Nudge effectiveness appears highly context dependent. Indeed, Thaler (2012) and Sunstein (BBC Radio 4, 2013) acknowledge that interventions that worked in one locality may not work in another. Even more, much of the nudging literature reports on experimental settings, based on relatively small and often homogenous

samples (i.e. a small group of students within a similar school or university to test the impact of a financial token reward). However, it remains at question whether the positive findings from such experiments can be truly extrapolated to hypothetical large-scale governance settings. It is a far extrapolation from, say, an experiment that shows that people are willing to carry out a simple task for a token reward, such as solving a crossword puzzle for \$5 (Kamenica, 2012), to expecting that rewarding teenagers with \$1 a day will make them reconsider getting pregnant (a nudge proposed in Thaler and Sunstein, 2009). In other words, policymakers, practitioners and academics interested in nudging need to be careful when reading sweeping generalizations about nudging as a governance intervention – whether these are about opportunities *or* the risks that nudging may bring.

Also, nudging trials and experiments have, so points out the literature reviewed, predominantly been carried out to test if nudging may assist in achieving desired collective ends (i.e. these trials and experiments are in line with what we refer to as type-2 nudging). As some critically state, in such examples the intervention is more like a ‘shove’ than a nudge (see Truog, 2012). This may give more food for thought to those who fear that nudging is nothing more than another approach of governments to take away individuals’ freedom. Also, the nudges we have traced in the literature predominantly are additions to existing and more traditional governance interventions, such as taxes or direct regulation. This may raise the question whether nudging truly is a fully fledged and independent governance strategy, or whether it is nothing more than an approach for revamping existing (and poor performing) governance interventions. As other scholars have critically stated, ‘[n]udges may only make limited difference and focus on them may take away from more difficult but more effective reforms’ (Rainford and Tinkler, 2011: 13).

Finally, nudging evaluations have thus far not compared nudging as a governance strategy or instrument to other governance strategies or instruments. In the range of experiments, trials and settings reviewed (see Table 1), a nudge was *added* to an existing governance instrument or a set of existing governance instruments, but did not *replace* an existing governance instrument. As such, it remains unclear whether nudging has a comparative advantage over other governance strategies or instruments (we traced only one scholarly work that actually compares the performance of nudges with the performance of other governance instruments, see Galle, forthcoming).

### *A focus on hard and soft outcomes*

Although some scholars argue that the uncertainty of nudge effectiveness poses unacceptable risks (e.g. Bradbury et al., 2013; Goodwin, 2012), we hold the opinion that the baby should not be thrown out with the bathwater. The experimental nature of, for instance, the BIT fits a larger trend of experimentalist governance that seeks to improve regulatory and governance outcomes through localized experimentation (Sabel and Zeitlin, 2011). Nudge theory at least provides a novel framework for policymakers to develop innovative regulatory interventions to play the ‘regulatory game’ between regulator and regulatee, where the latter learns to roll with the punches of the former, requiring for ongoing regulatory adjustment (see De Bruijn et al., 2007).

When evaluating examples of nudging, scholars may therefore wish to not only address ‘hard outcomes’, such as the number of people that quit smoking, or the amount of energy saved. Capturing ‘soft outcomes’ may be just as important to evaluate the value of nudging (see Darnall and Sides, 2008; Lyon and Maxwell, 2007; Rogers and Weber, 2010). What are the exact lessons learned from the particular nudge? Have particular collaborations of actors formed that would normally not have? Has the nudging experiment provided insights into how particular policies may be improved, even without being a success in terms of hard outcomes? Also, what is the value of knowing that a particular approach (a particular ‘nudge’) to steering people’s behaviour does not

work in achieving hard outcomes? Answering such questions would enrich the current evaluations of nudging practice.

## Conclusion

In this article, we sought to answer four related questions about nudging as a theory and practice. To conclude this article, we will briefly answer these questions.

### *Whether and how is nudging evaluable as a theory and a practice?*

In as far as nudging may be considered a theory, it is both a political philosophy or ideology and a theory of governance practice. As a practice, we have argued, nudging is either one of two things: changing actors' choice architectures to ensure they do not behave against their own self-interest as they would have done without the nudge (type-1 nudge), or changing actors' choice architectures to achieve desired collective ends (type-2 nudge).

As a political philosophy or ideology it cannot, almost by definition, be evaluated. The idea of libertarian paternalism may align with academics and policymakers' epistemology, or it may not. As a theory of governance practice nudging should, in principle, be evaluable. However, the theory as presented by Thaler and Sunstein (2009) comes with considerable complications and lacunae that complicate evaluations. First, there is no clear definition of what makes for a nudge, and how nudging differs from other governance interventions. Second, in unpacking Thaler and Sunstein's theory it has become clear that they in fact address two types of nudges, referred to as type-1 and type-2 nudges. For both types, however, it remains unclear how these are truly different from any of the governance interventions that had already been identified in the governance literature before the publication of *Nudge* (e.g. Baldwin and Cave, 1999; Salamon, 2001). When seeking to evaluate nudging as a governance practice, scholars thus need to clearly define what they consider a nudge.

In evaluating nudging as a practice, we believe scholars preferably evaluate it against another governance strategy or instrument, since this would help to better understand whether nudges outperform other governance interventions. We have not found many studies that actually compare nudges with other governance interventions (an exception is Galle, forthcoming). This may partly have to do with the fact that most studies we reviewed focus on type-2 nudges, which are more likely to be additions to existing governance instruments than actual new instruments in themselves. In short, evaluations of type-2 nudges are more likely to compare the 'old' governance instrument(s) with the 'revamped' governance instrument(s) when the nudge is or the nudges are added to it.

### *To what extent is solid evidence of nudge success available?*

The current evidence of nudging success goes both ways, according to our findings in reviewing the literature. In some circumstances, nudging achieves its desired outcomes, while in others it does not. This rather unexciting insight holds for both type-1 and type-2 nudges. Nudging successes are often presented with many caveats, the most recurring being that the finding holds in a particular setting. Scholars carrying out nudging evaluations, tests and trials are often careful in generalizing beyond the cases they study (policymakers and academics with an ideological preference for nudging, however, are more lenient in generalizing from these studies). It also often takes some close reading in what a nudging success (or for that matter, failure) exactly implies. For example, a relative increase of students choosing vegetables over less-healthy alternatives when

vegetables are presented more attractively does not tell anything about the absolute effect of this nudge, let alone what it does not achieve. In one of the examples cited, a relative increase of 23 percent implied a mere 8.2 percent absolute increase, and although a nudge was in place, the majority of students (56%, down from 64.2%) still chose the unhealthy alternatives (Hanks et al., 2013). Therefore, we urge scholars evaluating, trialling or testing nudging to be very clear about what their findings exactly show and what they do not show. This will help policymakers to make better informed decisions about whether to implement a nudge.

Strikingly, and as already addressed in answering the previous question, there hardly seems to be evidence of nudges being more successful as governance interventions than other governance interventions, such as direct regulation, taxes, subsidies or information supply. A fruitful area for future research may be such comparative research, because this would give policymakers a broader understanding as to where and when nudging may achieve better results than applying another governance strategy.

### *What is to be considered a nudge success?*

Most of the studies we have reviewed aimed at evaluating the effectiveness of a nudge as compared to a situation without that nudge. Studies are, in general, less clear about the efficiency of a nudge (i.e. what were the monetary costs or administrative efforts required to achieve the nudge outcome?). Another intriguing insight is that although scholars are caught up in fierce polemics about the legitimacy, accountability and transparency of nudging as a governance instrument, we have not traced any works that actually evaluated these criteria of real-world nudges. In terms of measuring nudging outcomes and nudging success, more is at stake in understanding whether nudging may be a successful governance intervention than what current evaluations present. Governance evaluations have a rich history of looking at multiple evaluation criteria seeking to present a holistic understanding of the performance of governance interventions (e.g. Armstrong, 2005; Bovens, 1998; Bovens et al., 2008; Hodge, 2000; Lundqvist, 1988; Power, 1999; Verhoest et al., 2007). The current focus on nudging effectiveness only feels a bit thin, in our opinion.

Also, current evaluations seek to judge nudging on its ability to achieve 'hard outcomes'. A lack of (evidence of) 'hard outcomes' is normally used by those critical to nudging's political philosophy as a reason to write off this governance instrument. Yet, by looking at the 'soft outcomes' of real-world nudges, valuable lessons may be learned.

### *What data and evaluative techniques may assist in evaluating Nudge theory beyond individual cases?*

With nudging being such a novel terrain in real-world governance, we believe that well-crafted comparative studies are necessary to better understand the performance of nudges as a governance intervention. Experiments from the behavioural sciences provide fruitful ground as starting points for hypotheses; however, it remains to be seen if the findings from lab-experiments and quasi-experiments are replicated in real world settings. Also, with nudging being such novel terrain it is unlikely that large-scale cross-sector, cross-time and cross-country studies can already be carried out. We expect more from explorative medium-n studies which seek to understand how a particular nudge operates in a particular set of contexts (Goertz and Mahony, 2012). In such studies, scholars may seek to understand the relative and absolute performance of the nudge (as compared to another governance intervention or a situation without the nudge, as well as compared to the goal of the nudge), and they may seek to understand whether a nudge achieves sweeping results in achieving desired collective ends (type-2 nudge), or in achieving

that people are healthier, wealthier and happier than they were without the nudge in place (type-1 nudge, following Thaler and Sunstein, 2009).

We also believe, and once more stress, that scholars interested in evaluating nudges need to be very careful in what to consider a nudge. For instance, some ‘nudging’ research lauds the finding that overall donations increase if this desired behaviour is somehow recognized (‘recognition’ is the nudge in this example; John and Stoker, 2010). Yet, the broader literature on donations, and philanthropy more generally, has for long been aware that recognition is but *one* of the many mechanisms that drive charitable giving (for a review of the literature, see Bekkers and Wiepking, 2010). We consider nudging not so much a revolution, but a logical evolution in the governance landscape (see Bradbury et al., 2013) – a landscape that has been subject to continuous change (see Baldwin et al., 2011; Levi-Faur, 2011; Sparrow, 2000). As we have discussed in the introduction, nudging has rapidly developed from a governance mechanism to a governance virtue. It is now time to evaluate whether the mechanism is worthy of this status as a virtue.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Notes

1. In this article, we work with the 2009 paperback edition of the book (Thaler and Sunstein, 2009).
2. Cass Sunstein joined the White House as Administrator of the White House Office of Information and Regulatory Affairs after publishing *Nudge*. Here his focus was on better regulation, not the application of behavioural insights per se. Yet, President Obama’s interest in nudging has been well-documented (Legett, 2014).
3. The Obama Administration document referring to the possible instalment of a Behavioral Insight Team was first made public by FoxNews, see: <http://www.foxnews.com/politics/interactive/2013/07/30/behavioral-insights-team-document/> (accessed 22 August 2013)

## References

- Ariely D (2008) *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: HarperCollins.
- Armstrong E (2005) *Integrity, Transparency and Accountability in Public Administration: Recent Trends, Regional and International Developments and Emerging Issues*. New York: United Nations.
- Baldwin R and Cave M (1999) *Understanding Regulation. Theory, Strategy and Practice*. New York: Oxford University Press.
- Baldwin R, Cave M and Lodge M (2011) *Understanding Regulation: Theory, Strategy and Practice*, 2nd edn. Oxford: Oxford University Press.
- Bardach E and Kagan RA (1982) *Going by the Book: The Problem of Regulatory Unreasonableness*. Philadelphia, PA: Temple University Press.
- BBC Radio 4 (2013) Analysis: Nudge theory in practice. *Current Affairs*. Available at: <http://podbay.fm/show/267300806/e/1364245200> (accessed 22 August 2013).
- Bekkers R and Wiepking P (2010) A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly* 40(5): 924–973.
- Bell S and Hindmoor A (2009) *Rethinking Governance*. Cambridge: Cambridge University Press.
- BIT (2013) Behavioural insights team. Available at: <https://www.gov.uk/government/organisations/behavioural-insights-team> (accessed 22 August 2013).
- Bovens M (1998) *The Quest for Responsibility*. Cambridge: Cambridge University Press.
- Bovens M (2010) Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics* 33(5): 946–67.

- Bovens M, Schillemans T and 't Hart P (2008) Does public accountability work? An assessment tool. *Public Administration* 86(1): 225–42.
- Bradbury A, McGimpsey I and Santori D (2013) Revising rationality. *Journal of Education Policy* 28(2): 247–67.
- Broder J (2012) Powerful shaper of U.S. rules quits, with critics in wake. *New York Times*, 4 August 2012, p. A1.
- Brown P (2012) A nudge in the right direction? *Social Policy & Society* 11(3): 305–17.
- Cabinet Office (2012) *Applying Behavioural Insights to Reduce Fraud, Error and Debt*. London: Cabinet Office.
- Cabinet Office (2013a) *Applying Behavioral Insights to Charitable Giving*. London: Cabinet Office Behavioural Insights Team.
- Cabinet Office (2013b) *Applying Behavioural Insights to Organ Donation*. London: Cabinet Office Behavioural Insights Team.
- Chhotray V and Stoker G (2010) *Governance Theory and Practice*. Houndmills: Palgrave.
- Cialdini R (2009) *Influence: The Power of Persuasion*. New York: William Morrow.
- Croci E (2005) *The Handbook of Environmental Voluntary Agreements*. Dordrecht: Springer.
- Darnall N and Sides S (2008) Assessing the performance of voluntary environmental programs: Does certification matter? *Policy Studies Journal* 36(1): 95–117.
- De Bruijn H, Ten Heuvelhof E and Koopmans M (2007) *Law Enforcement: The Game Between Inspectors and Inspectees*. Boca Raton: Universal Publishers.
- Dorey P (2014) *Policy Making in Britain: An Introduction*, 2nd edn. London: SAGE.
- European Commission (2013) Behavioural economics. Available at: <http://is.jrc.ec.europa.eu/pages/BE/BEindex.html> (accessed 22 August 2013).
- Fox News (2013) Gov't knows best? Available at: <http://www.foxnews.com/politics/2013/07/30/govt-knows-best-white-house-creates-nudge-squad-to-shape-behavior/> (accessed 22 August 2013).
- Gaker D, Zheng Y and Walker J (2010) Experimental economics in transportation. *Transportation Research Record* 2156(1): 47–55.
- Gallagher R (2012) *Applying Behavioural Insights to Reduce Fraud, Error and Debt*. London: British Academy.
- Galle B (forthcoming) Tax, command ...or nudge? Evaluating the new regulation. *Texas Law Review*.
- Gneezy U and Rustichini A (2000) Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3): 791–810.
- Goertz G and Mahony J (2012) *A Tale of Two Cultures*. Princeton, NJ: Princeton University Press.
- Goodwin T (2012) Why we should reject 'nudge'. *Politics* 32(2): 85–92.
- Halpern D (Producer) (2012) Applying behavioural insights and changing how Whitehall thinks and acts. *Nudge and Beyond: Behavioural Science, Policy and Knowing What Works*. Available at: <http://www.youtube.com/watch?v=E-CXdFKZIMk>.
- Hanks A, Just D and Wansink B (2013) Smarter lunchrooms can address new school lunchroom guidelines and childhood obesity. *The Journal of Pediatrics* 162(4): 867–69.
- Hausman D and Welch B (2010) To nudge or not to nudge. *Journal of Political Philosophy* 18(1): 123–36.
- Haynes L, Service O, Goldacre B and Togerson D (2012) *Test, Learn, Adapt*. London: Cabinet Office.
- Hodge GA (2000) *Privatization: An International Review of Performance*. Boulder, CO: Westview Press.
- John P and Richardson L (2012) *Nudging Citizens Towards Localism?* London: British Academy.
- John P and Stoker G (2010) *How Experiments Can Help Get Britain to the Big Society*. Available at: <http://www.civicbehaviour.org.uk/>.
- Johnson E and Goldstein D (2003) Do defaults save lives? *Science* 302(5649): 1338–9.
- Jolls C, Sunstein C and Thaler R (1998) A behavioral approach to law and economics. *Stanford Law Review* 50(5): 1471–550.
- Kahneman D (2011) *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kamenica E (2012) Behavioural economics and psychology of incentives. *Annual Review of Economics* 4(1): 427–52.
- Keizer K (2012) *The Spreading of (Dis)order*. London: British Academy.



- Legett W (2014) The politics of behaviour change: Nudge, neoliberalism and the state. *Policy and Politics* 42(1): 3–19.
- Leonard T (2008) Book review: Richard H. Thaler, Cass R. Sunstein, *Nudge*. *Constitutional Political Economy* 19(1): 356–60.
- Levi-Faur D (2011) *Handbook on the Politics of Regulation*. Cheltenham: Edward Elgar.
- Lipkus I, Klein W, Sugg Skinner C and Rimer B (2005) Breast cancer risk perceptions and breast cancer worry: What predicts what? *Journal of Risk Research* 8(5): 439–52.
- Lundqvist LJ (1988) Corporatist implementation and legitimacy: The case of privatisation in Swedish public housing. *Housing Studies* 3(3): 172–82.
- Lyon TP and Maxwell JW (2000) Voluntary approaches to environmental regulation: An overview. In: Franzini M and Nicita A (eds) *Economic Institutions and Environmental Policy*. Aldershot: Ashgate, 142–74.
- Lyon TP and Maxwell JW (2007) Environmental public voluntary programs reconsidered. *The Policy Studies Journal* 35(4): 723–50.
- Michie S and West R (2013) Behaviour change theory and evidence. *Health Psychology Review* 7(1): 1–22.
- Nomura H, John P and Cotterill S (2011) The use of feedback to enhance environmental outcomes: A randomised controlled trial of a food waste scheme. *Local Environment* 16(7): 637–53.
- Osborne D and Gaebler T (1992) *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Reading: Addison-Wesley Publishers.
- Pinch T (2010) Comment on ‘Nudges and cultural variance’. *Knowledge Technology Policy* 23(3): 487–90.
- Pink D (Producer) (2009) The puzzle of motivation. [Video] Available at: [http://www.ted.com/talks/dan\\_pink\\_on\\_motivation.html](http://www.ted.com/talks/dan_pink_on_motivation.html) (accessed 19 December 2013).
- Pollit C (1995) Justification by works or by faith? Evaluating the new public management. *Evaluation* 1(2): 133–54.
- Power M (1999) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Rainford P and Tinkler J (2011) *Designing for Nudge Effects*. London: LSE Public Policy Group.
- Richards D, Smith MJ and Hay C (eds) (2014) *Institutional Crisis in 21st Century Britain (Understanding Governance)*. Houndsmills: Palgrave MacMillan.
- Rogers E and Weber E (2010) Thinking harder about outcomes for collaborative governance arrangements. *The American Review of Public Administration* 40(5): 546–67.
- Rose J (2011) Diverse perspectives on the Groupthink theory. *Emerging Leadership Journeys* 4(1): 37–57.
- Sabel C and Zeitlin J (2011) Experimentalist governance. In: Levi-Faur D (ed.) *The Oxford Handbook of Governance*. Oxford: Oxford University Press, 169–85.
- Salamon L (ed.) (2001) *The Tools of Government: A Guide to the New Governance*. Oxford: Oxford University Press.
- Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ and Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18(5): 429–34.
- Science and Technology Committee (2011) *Behaviour Change: Second Report*. London: Authority of the House of Lords.
- Science and Technology Committee (2014) *Letter to the House of Lords*, 22 July 2014. London: The Science and Technology Select Committee. Available at: <http://www.parliament.uk/documents/lords-committees/science-technology/behaviourchange/behaviourchange/Ltr20140722.pdf>.
- Selinger E and Whyte K (2011) Is there a right way to nudge? *Sociology Compass* 5(10): 923–935.
- Shariff A and Norenzayan A (2007) God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science* 18(9): 803–9.
- Simon HA (1945) *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*. New York: Free Press.
- Sparrow MK (2000) *The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance*. Washington, DC: Brookings Institution.
- Suptoi A (2007) *Homo Juridicus on the Anthropological Function of the Law*. London: Verso.
- Thaler R (2012) Smart disclosure: The regulatory future. Available at: <http://www.youtube.com/watch?v=S5z2e0xriCU> (accessed 22 August 2013).

- Thaler R and Sunstein C (2009) *Nudge* (revised edition). London: Penguin.
- Truog R (2012) When does a nudge become a shove in seeking consent for organ donation? *The American Journal of Bioethics* 12(2): 42–4.
- Van der Heijden J (2013) Looking forward and sideways: Trajectories of new governance theory. *Amsterdam Law School Research Paper* 04: 1–26.
- Van der Heijden J and Ten Heuvelhof E (2012) The mechanics of virtue: Lessons on public participation from implementing the Water Framework Directive in the Netherlands. *Environmental Policy and Governance* 22(3): 177–88.
- Verhoest K, Verschuere B and Bouckaert G (2007) Pressure, legitimacy, and innovative behavior by public organizations. *Governance: An International Journal of Policy, Administration and Institutions* 20(3): 469–97.
- White M (2013) *The Manipulation of Choice: Ethics and Libertarian Paternalism*. Hampshire: Palgrave Macmillan.
- Whitehead M, Rhys J, Pykett J and Welsh M (2012) Geography, libertarian paternalism, and neuro-politics in the UK. *The Geographical Journal* 178(4): 302–7.
- Willis L (2013) When nudges fail: Slippery defaults. *University of Chicago Law Review* 80(3): 1155–229.

Mark Kosters has recently worked in the Australian Government's Office of Best Practice Regulation and Deregulation Policy Division.

Jeroen van der Heijden is an associate professor at the Australian National University and the University of Amsterdam. His research interest is in regulation and governance.