Nudging Cannot Solve Complex Policy Problems

Evan Selinger and Kyle Powys Whyte***

We deepen Adam Burgess' insight that under current conditions nudging cannot solve complex policy problems reliably and without controversy. We do so by integrating his concerns about nudging into Braden Allenby and Daniel Sarewitz's three-leveled model of the basic problems technology can address and generate. We use this model to explain why the UK experiment with nudging has revolved around techno-fixes with limited policy potential, and conclude that nudging is best seen as an emerging form of soft law.

I. Introduction

In theory, nudging can produce behavioral change people value. But under current conditions, nudging cannot solve complex policy problems. Adam Burgess[1] provides an excellent account of why some advocacy of nudges tends to be overblown. We will deepen Burgess' sociological and historical analysis by integrating his insights into a model of the basic problems that technology can address and generate, which we draw from systems engineering, philosophy, and science and technology studies. The model clarifies why the UK experiment with nudging has revolved around techno-fixes with limited policy potential. Before discussing the model, we will reconstruct some of Burgess' insights from the 2012 paper.

1. Burgess' concerns about nudging

a. Nudging is not the same as behavioral economics

Although nudging is associated with the 40 year-old field of behavioral economics, the two are not identical. Behavioral economics refers to a research field that has found human decision making to be less rational in some circumstances than classical economists typically assume. Nudges are attempts to change human behavior by using some of the findings of studies from behavioral economics. But choice architects must engage in the speculative and experimental process of creating the right interventions (e.g. technologies, displays of choices, built-environments, etc.) that are adjusted for the biases and improve undesirable behaviors on average. Burgess provides an interesting discussion of a point that has been featured in other contexts: Behavioral economics research can show why people might make certain decisions, but it is not robust enough to cover reliable predications about how people will behave in non-laboratory environments where variable perceptions of meaning exist.[2]

b. Proponents of nudging are overconfident

Nudging was branded in the United States as a "third way" between state interventionism and market driven norms.[3] David Cameron and the Conservative Party subsequently claimed nudging could be a form of "MINDSPACE" governance that helps usher

* Associate Professor, Rochester Institute of Technology.

** Assistant Professor, Michigan State University.

1 Adam Burgess, "'Nudging' Healthy Lifestyles: The UK Experiments with the Behavioral Alternative to Regulation and the Market", 3 *EJRR* (2012), this issue.

2 See also Evan Selinger and Kyle Powys Whyte, "Competence and Trust in Choice Architecture", 23 *Knowledge, Technology & Policy* (2010), pp. 461–482.; Kyle Powys Whyte, Evan Selinger, Arthur Capan et al., "Nudge, Nudge or Shove, Shove-The Right Way for Nudges to Increase the Supply of Donated Cadavor Organs", 12 *American Journal of Bioethics* (2012), pp. 32–39; Evan Selinger and Kyle Whyte, "Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture", 5 *Sociology Compass* (2011), pp. 923–935.

3 Anuj C. Desai, "Libertarian Paternalism, Externalities, and the 'Spirit of Liberty': How Thaler and Sunstein Are Nudging Us toward an 'Overlapping Consensus'", 36 *Law & Social Inquiry* (2011), pp. 263–295.

https://doi.org/10.1017/S186729X0000177X Published online by Cambridge University Press

in the "cultural revolution" of the "big society" (and its ethos of personal responsibility) and invested in the "Behavioral Insight Team" (so-called "Nudge Unit") "at a time when everyone else in the administration is experiencing significant cuts to personnel within the Prime Minister's Cabinet Office" . The Nudge Unit proclaimed that "There is no reason why we cannot succeed in tackling today's rising tide of chronic lifestyle-related disease." This hype is ironic given that health problems can involve value-laden disputes over ideals, outcomes, and even the parameters of rationality. Indeed, health problems are often framed differently (even incommensurably), giving rise to numerous conflicts, such as those over the best assessment metrics. They also can involve cause-and-effect relations that are so hard to decode that seemingly reasonable attempts at solving problems end up making things worse through unintended consequences. By themselves, behavioural solutions like nudges can be too reductive to cope with such complexity. Some critics thus worry that overselling the potential of nudges created opportunity costs: "Other critics have attacked the British application of nudging on the grounds that it wrongly precludes more regulation and legislation to promote health." Moreover, nudging may not yield promised savings. A main reason why British citizens initially found the idea of nudging appealing is that they were told it would help the government do more for society than expensive alternatives. But evidence of savings is lacking, and given the limited knowledge base, it was unrealistic to expect strong savings to occur early on.

c. Nudging is limited by conflicted ideals about freedom

Nudging tends to work best when users are unaware that their behavior is influenced by choice architecture. In order for citizens to be comfortable with nudges, two conditions thus need to be satisfied: first, citizens need to be comfortable having their behavior influenced at levels below the threshold of conscious awareness; second, citizens need to perceive nudge designers as trustworthy. According to Burgess, these conditions are difficult to meet because British culture is fundamentally conflicted about the value of autonomy. In fact, concern about the government using nudges for Orwellian purposes catalyzed critics to evoke Enlightenment style criticisms of manipulation.

d. Nudging may infantilize

Many believe that moral maturity requires individuals who can grasp the difference between right and wrong and self-direct towards righteous ends in the face of the temptations pulling in other directions. Because nudging steers user behavior without teaching them anything or improving their skills, populating policy with nudges runs the risk of infantilizing citizens – treating them like children to be controlled by others.[4]

e. Nudging is technocratic

Advocates represent nudging as a means to help individuals achieve goals they set for themselves. Nudging is thus formally differentiated from propagandizing. Nudges are in principle compatible with democratic politics because they simply realize the will of the people. Nevertheless, nudging relies on experts to determine "solutions to assumed problems," while preventing "wider issues," such as politically based arguments against a behaviour being legitimately classified as socially undesirable, from figuring in "its landscape." For example, "Schemes such as to reduce the amount of sex university students may be having are explored without hesitation, suggesting how marginalized concerns for privacy and liberty have become. It would appear to confirm that a managerial approach to society is firmly established, and manipulating information for the 'greater good' of avoiding risk or improving behavior unremarkable and assumed to be unproblematic." These are the sorts of decisions that people are not always comfortable having experts and bureaucrats make for them.

2. The fundamental problem of nudging policy

Burgess' insights revolve around a major concern: Under current conditions, nudging cannot solve complex policy problems reliably and peacefully. In this

4 Luc Bovens, "The Ethics of Nudge", in Till Yanoff and Sven Ove Hansson (eds.), *Preference Change*, (New York: Springer, 2009).

https://doi.org/10.1017/S1867299X00001777X Published online by Cambridge University Press

section, we develop the insight further by integrating Burgess' views into a model advanced by Braden Allenby and Daniel Sarewitz (2011)[5] that clarifies the basic problems technology can address and the main problems its use can generate. We are focusing on technology because policy applications of nudging tend to be instantiated as changes in technological or built-environment design that adhere to the logic of "techno-fixes" – a concept elaborated upon below.

The Allenby-Sarewitz model has three levels. Level I is "shop-floor" technology embedded in a "simple" "physical system." In this domain, techno-fixes are appropriate solutions for narrowly circumscribed problems: "Most of the necessary relationships amongst goals, means, and causality have already been captured in a physical system that can be used with confidence that a given input will produce a desired output."[6] Examples include vaccines, cars, and highly developed neuropharmaceuticals that can be used to reliably achieve goals as basic as staying protected from diseases, driving on a road in a desired direction, and performing better at a mental task (ibid).

Level I technologies can be designed to "eliminate risk" or transform environments so that they contain "less risky alternatives" than otherwise would be present.[7] Approaching the "hierarchy of control" in this way can differ significantly from adopting alternative approaches to risk management that "rely on institutional and social systems."[8] Allenby and Sarewitz offer the following example as an illustration of the divergent approaches:

> Consider...the use of toxic chemicals in a manufacturing process. One way to manage such a risk to employees is to provide appropriate protective equipment and mandate its use. But protective equipment can break down, be misplaced, or simply not used. (Employees often dislike protective equipment because it is uncomfortable and makes their job harder; managers may not be as strict as they should be in requiring it as they should be, because it can impeded productivity.) Alternatively, one can redesign the manufacturing process to use less toxic materials in the first place. The former ... is never as effective as the latter (a "vaccine approach").[9]

On the basis of comparing the options available in this case as well as a host of similar ones, they conclude that, "The more responsibility for safety you can transfer from human and institutional decisions processes to technologies themselves, the safer the system will be, all else equal."[10] That is the beauty of a techno-fix: means and ends are unified.

If all major problems could be solved by techno-fixes, life might be easier. But, techno-fixes have a limited range of application precisely because social life is immensely complex.

This is where Level II comes in. It consists of "complex socio-technical systems" that are "infinitely less predictable" than Level I technology, but which support it through delivery, coordination, and regulatory systems.[11] For example, airplanes are a Level I technology, but "airline corporations, the government security apparatus as applied to air travel, and market capitalism in route pricing" are Level II subsystems.[12] While airplanes cannot be integrated into society without Level II support, mismatches between Level I and Level II operations frequently arise that cause travelers to experience immense frustration: good safety records co-exist with high incidents of delays and cancellations and poor customer service. Frustration is not the only problem that can occur when Level II issues are at stake. In the last section, we reviewed Burgess' analysis of how concern over governmental development and administration systems impede citizen enthusiasm for using nudging as a policy tool.

Let us return to the vaccine example (Level I technology), but consider it from the perspective of Level II activity related to risk management and public health. Malaria is a serious problem in Africa and currently no vaccine exists. The most useful preventive technology is bed netting. Despite heavy expense and effort, by 2005 "only 3 percent of African children in malarial regions were sleeping under insecticide-impregnated bed nets."[13] Policy consensus emerged that problems concerning local context – mismatches between the technology's physical properties and social/cultural norms – created the

5 Braden Allenby and Daniel Sarewitz, *The Techno-Human Condition* (Cambridge, MA: MIT Press, 2011).

6 *Ibid.*, p. 107.

7 *Ibid.*, p. 52.

8 *Ibid.*

9 *Ibid.*, pp. 51–52.

10 *Ibid.*, p. 52.

11 *Ibid.*, p. 38.

12 *Ibid.*

13 *Ibid.*, p. 48.

https://doi.org/10.1017/S1867299X00001177X Published online by Cambridge University Press

Table 1

Level I Nudging	Many nudge proposals have not qualified as Level I interventions when embodied in artifacts and built-environments. Currently, the cause-and-effect relations are not nearly as predictable as vaccines and other paradigmatic Level I technologies.
Level II Nudging	Deploying nudges has been a highly problematic endeavor because, in addition to the causal problem, the systems required to fund their design, administer them, prime the public for widespread use, and convince the public that nudges are innovative tools that enhance well-being across society, are plagued by ethical and political concerns related to democratic governance, cognitive liberty, fair use of public funds, and opportunity costs. Unlike the IVM example, technical specifications are misaligned with social and cultural norms/values.
Level III Nudging	Nudges are not really discussed in this register. For example, although discussions abound about how nudging can reduce energy consumption, nudge proponents do not even attempt to discuss at all the complex interventions that are needed to shift diverse citizens in diverse parts of the world away from a CO_2 intensive society or adapt to the changes that might accompany the perpetuation of CO_2 intensive industries, infrastructures, and lifestyles.

biggest obstacles to adoption. To make progress, an approach called "Integrated Vector Management" (IVM) was embraced that tailored interventions to local factors, including "type of mosquito and malaria parasite, climate, cost, and available resources."[14]

While this initiative proved exceptionally effective, Allenby and Sarewitz note that, "The main selling point of IVM – its sensitivity to context – is also its weakness."[15] By weakness, they mean adjusting interventions to context can be far less reliable and far more resource intensive than standardized techniques that allow the same action to invariably yield the same result. For this reason, even if a highly reliable, standardized malaria vaccine were invented, it would be unrealistic to expect that gains in a Level I system would scale-up smoothly to the Level II action of administering to human patients: "There would, no doubt, be controversies over appropriate policies for vaccine delivery, there would be organizational dysfunction at various levels, and a continued lack of infrastructure for delivering health care in parts of Africa."[16] In other words, problems would arise that are structurally similar to the issues when employees are mandated to protect themselves by wearing safety equipment.

Level III technology is the most difficult to understand. Fortunately, not much needs to be said about it here, as the problems with nudging mostly concern misalignment between Levels I and II. Simply put, Level III concerns "wicked problems" occurring at a "whole earth level" in which multiple systems, operating with different norms and at different scales, interact with one another to produce emergent behavior that can be exceptionally difficult, if not impossible, to predict. For example, if an airplane functions at Level I, and air traffic controllers act at Level II, then the global problem of infectious disease travelling rapidly due to the pervasiveness of international flights can be considered a Level III effect. The complexity at issue here strains moral imagination and requires "macro-ethical" strategies to cope with.[17]

Although the model is descriptive, it can be applied to further two important normative ends.

1. It can help us detect invalid predictions that are made about projects revolving around artifacts and technological systems. For example, people who believe that inventing a new solar technology will solve the energy crisis confuse Level I and Level III systems. Likewise, people who claim that meaningful economic development will result from transferring a vaccine to a developing country mistake Level I and Level II systems.
2. It can help us to determine when the right and wrong scale of assessment is used to judge a technical project or a technical system. In the transhumanism debates, for example, Allenby and Sarewitz believe confusion arises because proponents address Level I while opponents address Level II and Level III.

14 *Ibid.*, p. 48–49.

15 *Ibid.*, p. 50.

16 *Ibid.*, p. 49.

17 *Ibid.*, p. 184.

https://doi.org/10.1017/S1867299X0000177X Published online by Cambridge University Press

Let us apply the model to the situation of using nudges in British policy, accepting all the information Burgess provides as accurate.

Looking at things through Allenby and Sarewitz's three levels, the hype surrounding the anticipated behavioral results and savings of nudging can be seen as misguided. Not only are the robust cause-effect relations that typify Level I absent in nudges, but the ethical and political concerns expressed at Level II should have been foreseen and better dealt with. Having been subject to critique, the Behavioral Insight Team is now more sensitive to these issues. In a recently published report[18] they stated that more small scale experimentation needs to take place in order to assess the feasibility of certain nudges (and other behavioral interventions of interest). But these experiments only promise to shed light on the Level I effects of nudges. That is, they can only help us understand whether plausible nudge ideas could work relatively well in certain contexts. They do not appear to address the Level II problems that Burgess identifies. Approaches borrowing from stakeholder dialog models found in participatory governance are needed to address these concerns. But even these models may be insufficient, in-themselves, to guide policy makers through the difficulties that can arise when gains in contextual-sensitivity come at the expense of reliability and increase an operation's costs.

II. Conclusion

Was solid evidence ever really offered that nudging could contribute significantly to policy? Or, was the policy rhetoric always inflated? It should come as no surprise to the reader, but our respective answers are: no and yes.

While Burgess focuses on UK problems, the fact is mistakes have been made from the start concerning discussion of nudges. These mistakes began in the United States in the very text that got the ball rolling: *Nudge*. One of Thaler and Sunstein's favored examples is the "Toxic Release Inventory."[19] It is a list of "potentially hazardous chemicals that have been stored or released into the environment" archived on the U.S. Environmental Protection Agency's website.[20] Thaler and Sunstein depict the information as potentially "the most unambiguous success story in all of environmental law."[21] They give such high praise because the data created an "environmental blacklist" that generated bad publicity for offending companies; in some cases, adverse reaction led to stock price devaluation. [22]

Unfortunately, as commentators have noted, the Toxic Release Inventory is not a nudge.[23] Nudges, as per Thaler and Sunstein's definition, cannot change behavior by significantly altering economic incentives. The inventory, however, brings about environmental compliance by virtue of increasing polluting costs.[24] It is surprising that Thaler and Sunstein are silent on this. Indeed, their analysis actually emphasizes the inventory's ability to cause economic harm: "If companies are able to reduce emissions at a low cost, they will do so, simply in order to avoid the bad publicity and resulting harms."[25] In fact, not only do Thaler and Sunstein mistakenly describe the inventory as a nudge, but they also suggest an analogous version – also erroneously depicted as a nudge – should be created to reduce greenhouse gas emissions: "The government should create a Greenhouse Gas Inventory (GGI), requiring disclosure by the most significant emitters." [26]

Another much discussed policy nudge that Thaler and Sunstein offer is the use of mandated choice schemes to increase the supply of donated organs. This policy has recently been instituted in Great Britain as a "required choice" policy.[27] They contend that mandating people to choose whether or not to donate their organs as a requirement for obtaining a driver's license would increase the amount of registered donors, while "imposing essentially no new burdens on taxpayers," and avoiding the political barriers of presumed consent."[28] Unfortunately, this example is not persuasive, either. It has been argued that Thaler and Sunstein's view of how this nudge would

18 BIT, "Behavioral Insights Team Annual Report", Cabinet Office (2010-2011).

19 Richard Thaler and Cass Sunstein, *Nudge: improving decisions about health, wealth, and happiness* (New Haven, CT: Yale University Press, 2008), pp. 192–93.

20 *Ibid.*, p. 192.

21 *Ibid.*

22 *Ibid.*, p. 192–93.

23 Daniel M. Hausman and Brynn Welch, "Debate: To nudge or not to nudge", 18 *Journal of Political Philosophy* (2010), pp. 123–136.

24 *Ibid.*, p. 124.

25 Thaler and Sunstein, *Nudge*, *supra* note 19, at p. 193.

26 *Ibid.*

27 BIT, Behavioral Insights Team Annual Report, supra note 18

28 Thaler and Sunstein, *Nudge*, *supra* note 19, at p. 178.

https://doi.org/10.1017/S1867299X00001777X Published online by Cambridge University Press

motivate people to embody altruistic convictions is flawed.[29] Simply put, they overlook the fact that people can be influenced by perceptions of features found in and associated with the Department of Motor Vehicles. These features can be emotional triggers that "shove" people looking to obtain licenses toward opting against donation. Whereas the Toxic Release Inventory fails on definitional grounds, this one fails on empirical ones.

Now, Thaler and Sunstein do offer an example of a policy nudge that meets both definitional and empirical criteria. This example concerns Lake Shore Drive, a U.S. roadway that has stunning views of Chicago's skyline. One particular segment includes a series of S curves that require drivers to slow down to 25 mph. Distracted by scenery or failing to calculate depth properly, drivers often ignore the reduce speed limit sign. To alter the dangerous behavior, a visual illusion, induced through a series of painted white stripes, has been installed; the stripes incline drivers to slow down: "When the stripes first appear, they are evenly spaced, but as drivers reach the most dangerous portion of the curve, the stripes get closer together, giving the sensation that driving speed is increasing. One's natural instinct is to slow down."[30]

But if we assume this example works, the result actually does not bode well for the hype about nudges. Repainting the roadway is by no means a third alternative to regulation or the market. Road safety is the responsibility of federal and state regulators. The white stripes are simply the product of a decision by the traditional authorities. Second, the stripes do not actually eliminate the problem. According to a 2006 article, the Transportation Department "set aside $1.4 million for a two-year study that could result in a redesign of the Oak Street curve and improvements in accessing the lakefront."[31] There is still much more to be studied regarding the importance of wetness conditions, warning signs, and other environmental features. In this case, then, a nudge provides limited gains and, at best, is still not a Level I technology. Rather, nudging here is more like an emerging form of soft law that regulators can explore as an ordinary part of their charges.

On the basis of Burgess's concerns, our application of Allenby-Sarewitz model, and the problems identified with how *Nudge* treats policy, our conclusion is that under current conditions nudging cannot solve complex policy problems reliably and without controversy. The UK nudge experiment thus far primarily has treated nudges as techno-fixes that only can have limited policy value. Whether the Behavioral Insight Team makes the changes needed to improve Level I and II performance, or whether it is merely rhetorically gesturing to these improvements, remains to be seen. Nudges are best seen as an emerging form of soft law regulation.

29 Whyte, Selinger, Capan et al., "Nudge, Nudge or Shove, Shove", *supra* note 2.

30 Thaler and Sunstein, *Nudge*, *supra* note 19, at pp. 38–39.

31 Jon Hilkevitch, "Lake Shore Curve to Get More Alerts," Chicago Tribune, 24 July 2006.

https://doi.org/10.1017/S1867299X0000177X Published online by Cambridge University Press