

The Choice Architect's Trilemma

Chris Mills¹

Published online: 9 August 2017

© The Author(s) 2017. This article is an open access publication

Abstract Critics have long dismissed paternalistic choice architecture as conceptually muddled at best and oxymoronic at worst. In this article, I argue that this criticism remains true despite recent replies to the contrary. Further, I suggest that a similar conceptual criticism also applies to non-paternalistic choice architecture. This is due to a three-way tension between the effectiveness, avoidability, and distinctiveness of each nudge. To illustrate this tension, I provide a novel explanation of the mechanics of nudging and a taxonomy of these interventions. I then argue that choice architects who defend the distinctiveness of nudging according to how it guides our behaviour via our unreflective intuitive reasoning encounter a trilemma because the distinctiveness of nudging hinges on interventions being both avoidable and effective. Choice architects cannot achieve this aim without harnessing both our automatic and reflective systems of thought in tandem. However, this diminishes the distinctive character of nudging by bringing it closer to other traditional policy interventions. This establishes the choice architect's trilemma: a nudge is likely to be either ineffective, effective via some morally unacceptable means, or effective in a manner that is conceptually indistinct from other evidence-based policy interventions.

Keywords Choice architecture · Freedom of choice · Manipulation · Nudge · Paternalism

Cognitive psychologists and behavioural economists have long argued that our everyday decision-making is not a uniform process. Rather, our decisions result

✉ Chris Mills
c.mills.3@warwick.ac.uk

¹ School of Law, University of Warwick, Coventry CV4 7AL, UK

from the interplay between automatic and reflective aspects of our will.¹ The relationship between these systems of thought is more complex than we often acknowledge. We have far less awareness and control over our automatic reasoning than we commonly assume and yet this automatic reasoning is responsible for a far greater proportion of our behaviour than we usually realise (e.g. Kahneman and Tversky 2000; Kahneman 2003, 2011).

This research provides policy-makers with a clearer picture of how traditional policy interventions (such as incentive offers and coercive sanctions) change our behaviour. This picture allows us to improve these policies to better reflect how citizens actually make everyday decisions. However, this research has also informed the development of novel techniques of policy intervention. The traditional carrots and sticks offered by incentives and sanctions primarily appeal to our reflective reasoning by offering us reasons for and against certain outcomes. As this mode of thought is not the only one that we engage when making decisions, we can conceive of a different type of policy that primarily influences us via our automatic reasoning rather than our reflective reasoning. These policies will not provide us with explicit reasons for action. Rather, they will prompt our intuitive thinking in order to guide our behaviour. If these policies are effective then they are an important addition to the policy-maker's toolbox (Oliver 2013a).

One version of this form of policy-making has distinguished itself from its relatives in recent years by purporting to offer a distinctive and appealing combination of features. Defended most prominently by Richard Thaler and Cass Sunstein, this approach suggests that governments can nudge citizens towards making better unreflective choices without restricting their freedom of choice. According to Thaler and Sunstein, a nudge is any intervention to shape a decision-making situation '...that alters people's behaviour 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' (Thaler and Sunstein 2009, p. 6).²

Sunstein and Thaler offer two distinct reasons for why the state may nudge citizens' behaviour (Sunstein and Thaler 2003, p. 1193). First, the state may intervene in order to safeguard the well-being of the decision-maker. When the state intervenes to increase the likelihood that citizens succeed according to their own informed preferences, it acts on principles of *libertarian paternalism*.³ Alternatively, the state may intervene in order to secure the well-being of third parties. When the state intervenes to make it more likely that citizens will act to further the

¹ Stanovitch and West (2000) famously label these choice-determining phenomena 'System One' (automatic) and 'System Two' (reflective) processes.

² Sunstein and Thaler claim that a wide range of policy types fulfil this definition, including choice prompts, selective information provision, information framing, de-biasing, default rules, and more.

³ Camerer et al. (2003) name this approach *Asymmetric Paternalism*.

collective good or some other value of social justice, it acts according to principles of *libertarian benevolence*.⁴

To illustrate this distinction, consider a well-known example:

Cafeteria – Alice is put in charge of a cafeteria at her institution. She must make a multitude of decisions, including which foods to serve, which ingredients to use, and in what order to arrange the meal choices. After a few weeks in the job she notices that customers have a tendency to choose more of the items that are presented earlier in the line.⁵

In *Cafeteria*, Alice notices a significant framing effect: her customers' decisions are shaped by how she presents the food to them. If Alice adopts a principle of libertarian paternalism, she might present their food so that healthier or cheaper products appear earlier in the line. Either of these alterations will increase the likelihood that customers' decisions will unreflectively align with their own informed preferences toward those ends.⁶ Alternatively, Alice may adopt a principle of libertarian benevolence and present the food so that fair-trade products or products with a lower carbon footprint appear earlier in the line. Either of these alterations will likely make it easier for customers to act in a socially just manner.

The distinction between these two policy types should be clear. Both approaches conform to the basic principle of nudging—changing behaviour in a predictable fashion without restricting freedom of choice. For this reason, Sunstein and Thaler attach to them the prefix *libertarian*. The important difference between the two approaches lies in their stated aim. Libertarian benevolence is justified according to improvements in the well-being of third parties (or some other value of social justice) whereas libertarian paternalism is justified according to improvements in the welfare of the individual affected.⁷

This distinction shows us that nudges are not distinctive due to their aim of paternalising us or facilitating justice; most policy interventions do this. Rather, nudges are distinctive due to their method of bringing about these outcomes. Nudges distinguish themselves from other policy interventions by shaping our decisions *without* restricting our freedom of choice. Nudges do not offer any explicit incentives or threaten any significant costs. Rather, choice architects sidestep our reason-responsive faculties and design choice situations in order to exploit everyday

⁴ Korobkin (2009) names this approach *Libertarian Welfarism*. Sunstein and Thaler suggest that this approach justifies the state's attempts '...to promote benevolence, and to assist vulnerable people, without mandating behaviour in any way' (Sunstein and Thaler 2003, p. 1193). Although libertarian paternalism has received the lion's share of attention, some have explored the possibility and attractiveness of libertarian benevolence (Smith and McPherson 2009; Kelly 2013; Guala and Mittone 2015).

⁵ Adapted from Sunstein and Thaler (2003, p. 1164).

⁶ The truth of this claim depends on the preferences of Alice's customers. If the majority of her customers had informed preferences in favour of expensive unhealthy food, then Alice might present the food differently.

⁷ This contrast ensures that libertarian benevolence promotes a far broader range of values than its paternalistic sibling. As a result, these two justifications may conflict: policy interventions intended to improve individual well-being may create externalities that decrease social utility and harm the collective good. Conversely, policies that benefit the collective good may sacrifice the well-being of certain individuals.

‘Biases and Blunders’ (Thaler and Sunstein 2009, pp. 24–41) and ‘Behavioural Market Failures’ (Sunstein 2014, pp. 34–50) that characterise our automatic reasoning.⁸ As we can shape this form of reasoning without resorting to coercion, proponents believe that choice architecture is often preferable to traditional incentives and sanctions.

Debate over the permissibility of nudging has been hampered by a persistent and troubling concern that tracks a general scepticism that nudging is conceptually incoherent. Following Sunstein and Thaler, we can label this concern the *Oxymoron Objection*.⁹ This objection has persisted for many years, leading some to champion more conceptually robust behavioural interventions (Oliver 2013b, 2015; Grill 2014) and others to radically revise the nudge project in the hope of salvaging it (Saghai 2013). Indeed, this objection is so persistent that Sunstein (2015, p. 514) has come to accept the difficulty of general theoretical discussion of nudging, instead imploring proponents to avoid the ‘trap of abstraction’ and to assess the plausibility of each intervention on a case-by-case basis.

In this article, I explore the nature and scope of the oxymoron objection, explain why it persists, and assess its troubling consequences for choice architects. I begin by exploring the original form of the objection, which denies the compatibility between libertarianism and paternalism. I illustrate the failures of various rebuttals of this objection and argue that it continues to afflict paternalistic choice architecture. I then identify a broader novel form of the oxymoron objection that denies that nudging can simultaneously be effective, avoidable, and distinctive. I argue that the tension between these three characteristics supports the scepticism that even non-paternalistic nudges require further conceptual clarity. I conclude by assessing the difficulty that proponents face in resolving this trilemma and sketching out its likely consequences.

Is Libertarian Paternalism Still an Oxymoron?

In this section, I explore the original version of the oxymoron objection in order to identify the traditional source of conceptual confusion concerning nudges. This version of the objection is motivational as it concerns the compatibility of the libertarian elements of choice architecture with paternalism. According to this worry, the paternalistic justification for nudging is oxymoronic because interventions that aim to be libertarian cannot also be paternalistic.

There are two prominent responses to this objection. First, Sunstein and Thaler (2003, 2009, pp. 5–12) argue that there is nothing oxymoronic about achieving paternalistic ends via libertarian means. The libertarian state minimises the amount of coercion that it imposes on its citizens (e.g. Hayek [1960] 2006, pp. 11–20; Nozick 1974, pp. 26–28). Nudges are compatible with this aim because they are not

⁸ These aspects of our psyche include anchoring, attribute shrouding, the availability heuristic, framing, loss aversion, the representativeness heuristic, status quo bias, irrational temporal discounting, and unrealistic optimism.

⁹ See also Mitchell (2005).

coercive. Rather, they purport to improve our decisions without restricting our freedom of choice. Thus, nudging is compatible with a libertarian respect for freedom even if we nudge for paternalistic reasons.

This response faces a clear problem: libertarianism is not simply a claim about the means by which states act; it is also a claim about the sorts of policies that states choose to pursue. A principled stance against coercion is not the ultimate goal of libertarianism. Rather, the libertarian state protects certain core rights and freedoms of its citizens (e.g. rights of self-ownership). The concern with coercion that Sunstein and Thaler identify is symptomatic of this more fundamental cause because the need to protect core rights leads the libertarian state to avoid unnecessary coercion. So long as citizens can enjoy their rights while respecting the rights of others, then the libertarian state has little reason to intervene more intrusively in order to make further decisions on behalf of citizens. The libertarian state only intervenes to secure core rights and protect against their transgression.

This upstream focus on protecting core rights and freedoms ensures that paternalistic interference is unnecessary (and thus objectionable) to libertarians (Wall 2009).¹⁰ Libertarians believe that improvements in individual welfare are not sufficient justification for state intervention *regardless* of the means of that intervention. Libertarianism cannot consistently generate the reasons required to motivate paternalistic interference from its own basic principles. Therefore, paternalistic nudging cannot be justified on libertarian grounds. The libertarian state is a nightwatchman rather than a nanny.

Second, Sunstein (2014) provides an alternative defence of libertarian paternalism when he argues that the cognitive improvements that result from counteracting our blunders, biases, and behavioural market failures should fall within the legitimate scope of the libertarian state. These interventions improve choices in line with the wishes of decision-makers. Thus, if libertarians aim to ensure that citizens can enjoy their rights as they wish (consistent with the minimal infringement of the rights of others) then libertarian paternalism can improve the choices of citizens in line with this requirement.

He argues for this possibility through a pair of distinctions. First, we should distinguish between *hard* and *soft* paternalism (Sunstein 2014, p. 58). Hard paternalism promotes individual well-being by imposing *material* costs on individual decision-making. In contrast, soft paternalism imposes *non-material* costs on individual decision-making.¹¹ While this distinction will often be a matter of degree, fines and taxes are clear examples of the former while default options and information framing are obvious examples of the latter. Second, we should distinguish between *means* and *ends* paternalism (Sunstein 2014, pp. 61–71). Means paternalism facilitates our pursuit of our own conception of the good. In contrast, ends paternalism directly improves the content of our conception of the good.

¹⁰ For criticism concerning the likely anti-libertarian redistributive effects of nudging, see Mitchell (2005, pp. 1269–1275).

¹¹ Sunstein's use of this distinction differs from the traditional distinction between hard and soft paternalism coined by Joel Feinberg (1986, p. 15). Feinberg distinguishes between influencing voluntary and non-voluntary decisions.

These two distinctions complement each other: the first (between hard and soft paternalism) distinguishes between methods of interference while the second (between means and ends paternalism) distinguishes between the aims of interference. When evidence suggests that policy-makers could improve people's lives through *soft-means* paternalism, then libertarian paternalism suggests that they should do so. This permission is supposedly granted by the fact that soft-means paternalism allows individuals to more easily pursue their own conceptions of the good (means paternalism) without imposing significant material costs on themselves and others (soft paternalism). Paternalistic choice architecture should possess these characteristics, and interventions that possess these characteristics secure freedom of choice rather than undermine it. From this, Sunstein concludes that the very possibility of soft-means paternalism illustrates the compatibility between libertarianism and paternalism.

This response improves on the original rebuttal of the oxymoron objection because it speaks to *both* the aims and means of libertarianism. However, even if we assume that this response does succeed in linking libertarianism to paternalism where the prior argument fails, it still faces important objections from both anti-paternalists and paternalists.

On one hand, anti-paternalists may reject the distinction between ends and means paternalism as implausible. To see this, imagine that you are required to design a policy to improve the financial decisions of Alice and her co-worker Brian. Evidence suggests that you could design the architecture of a decision that both Alice and Brian regularly make so that it is more likely that both colleagues will save more of their money. This intervention helps miserly Alice because she values saving for its own sake. This intervention also helps risk-averse Brian because, although he does not value saving for its own sake, Brian does value saving as a means of protecting his future prospects against the risk of redundancy.

In this situation, saving is both an end in itself for Alice and, simultaneously, a means to some other end for Brian. This difference between Alice and Brian's conceptions of the good ensures that any intervention to change their saving habits will affect them differently. On a standard reading of the case, our interference with Alice's ends will incur a larger justificatory burden than our interference with Brian's means. This is because Alice's ends are more morally weighty than Brian's means (*ceteris paribus*).

Sunstein's move toward means paternalism acknowledges this important difference in the justificatory burdens generated by each intervention. Problems arise, however, when we acknowledge this moral fact in the light of the vast heterogeneity of conceptions of the good held by citizens residing in modern liberal societies. This heterogeneity makes it difficult for the state to know when a policy intrudes upon a decision concerning our ends or our means. Sunstein believes that principles of libertarian paternalism can justify the state intervening in order to improve Brian's life. Nevertheless, many policies designed to affect Brian's choices will also affect Alice's choices, and affecting Alice's choices is a more morally weighty endeavour to justify. Because of this, the state should act according to the belief that every soft-means paternalistic intervention may also be a form of soft-ends paternalism for some other person. Thus, the claim that the libertarian

paternalist state is permitted to pursue soft-means paternalism has little practical payoff in the face of epistemic uncertainty and social diversity.¹²

In light of this problem, anti-paternalists should insist that proving the mere conceptual possibility of means paternalism does not resolve the difficulty of isolating it from its more morally troubling sibling in real-world cases. Given the facts on the ground, proponents must change defensive tack and provide a normative justification of ends paternalism from deeper libertarian principles. As we have seen, this will require proponents to weigh up the moral harm of frustrating the rights of some citizens (like Alice) in order to facilitate other citizens (like Brian) in their enjoyment of their rights. This is the challenge awaiting those who seek to defend soft-means paternalism as a legitimate goal of libertarian policy-making.

Choice architects might respond to this challenge by appealing to the non-intrusive nature of soft-paternalism. Even if an intervention that is intended to be an instance of means paternalism backfires and transforms into an instance of ends paternalism, the intervention will not impose any significant material costs on those affected. Because of this, any resulting ends paternalism will not significantly restrict our freedom of choice and should be easily avoided. If Alice changes her mind and comes to doubt the value of saving as an end in itself, she will not be significantly frustrated in her pursuit of the good when her financial decisions are shaped as a side effect of the paternalism aimed at Brian.

Troublingly, this response fails to recognise that a constraint need not be weak just because it is non-material. The scorn of my peers may be far more damaging to me than any fine or financial penalty. As J. S. Mill famously notes, this scorn can leave us fewer means of escape and penetrate much more deeply into the details of our lives (Mill [1859] 1991, p. 9). This should lead us to conclude that Sunstein's defence of the soundness of libertarian paternalism cannot evade the oxymoron objection by relying on the possibility of soft-means paternalism. Anti-paternalists can object that relying on these distinctions underestimates the epistemic difficulties involved in targeting means paternalism in a diverse society, and the possible harshness of so-called soft-paternalism.

On the other hand, paternalists may challenge the paternalistic credentials of soft-means paternalism (e.g. Conly 2013; Arneson 2015). Accept for a moment that the above objection can be resolved and that we can plausibly distinguish between means and ends paternalism in a useful way. By doing so, we accept that a state may act for paternalistic reasons to help citizens live better lives. Paternalists may then ask whether such policies really are effective in achieving their paternalistic ends.

By leaving imprudent options available and instead relying on our (suitably influenced) cognitive biases to lead our decisions away from these options, nudges seem poorly suited to paternalistically improving our lives in many important cases.

¹² To minimise the risk of ends paternalism, the state must close the epistemic gap in order to better understand the structure of our preferences. However, this will be difficult for two reasons. First, it will require our preferences to be more transparent to observers than some suggest is possible (e.g. Rebonato 2012, pp. 153–163; White 2013, pp. 69–80). Second, it will require a large-scale invasion of privacy in order to collect the necessary information about our conceptions of the good (Kapsner and Sandfuchs 2015). Overcoming these difficulties will be an onerous task. Yet if choice architects cannot overcome these difficulties, the distinction between means paternalism and ends paternalism is of little practical use.

For example, irrational acts backed by strong motivations to act irrationally are likely to overcome nudges (Conly 2013, pp. 31–32). Sunstein's move towards soft-means paternalism exacerbates this concern. By reducing the scope of permissible nudging to soft-means paternalism, Sunstein effectively increases our opportunities for error and devalues nudging as an effective form of paternalism. Consequently, those with an appetite for effective paternalism are likely to look elsewhere for solutions to our poor decision-making (e.g. Conly 2013, pp. 29–33).

This gives us a second reason to think that Sunstein's response fails to rebut the oxymoron objection: even if it were true that we can isolate incidents of soft-means paternalism, such interventions may be too harsh and broad for anti-paternalists and too weak and narrow for paternalists.¹³ These objections show that the defence of libertarian paternalism as a conceptually coherent project still has some way left to go. No response to the oxymoron objection against libertarian paternalism can rely on the supposed compatibility between libertarian methods and paternalistic aims. Rather, it must explain how paternalistic aims can be squared with deeper libertarian principles in order to fully respond to the objection. Drawing a distinction between ends and means paternalism cannot achieve this goal because this distinction raises a pair of further objections. Either libertarian paternalism's reliance on means paternalism is implausible (because means paternalism is inseparable from ends paternalism in practice), or this distinction is plausible but the soft-means paternalism that it produces is unsavoury for both sides of the paternalism debate. As neither of these conclusions is an attractive one for libertarian paternalists to draw, we should conclude that the oxymoron objection continues to loom large over paternalistic choice architecture.

Understanding the Mechanics of Nudging

In the previous section, I identified and explored the traditional conceptual objection against nudging. That such a fundamental challenge to the coherence of libertarian paternalism continues to persist should raise alarms. Yet the persistence of this version of the oxymoron objection is not fatal to the overall nudge project so long as some nudges evade it. We saw at the outset that libertarian paternalism does not exhaust the possible justifications for nudging. This will lead some choice architects to look to libertarian benevolence in the hopes of escaping the oxymoron objection. In what follows, I argue that this strategy is misguided because we can identify a different version of the oxymoron objection that more broadly afflicts both paternalistic and non-paternalistic choice architecture. This mechanical objection concerns how nudges work rather than the compatibility of their conflicting motivating principles. It suggests that the mechanisms of psychological influence used by choice architects to shape our behaviour guarantee that most interventions cannot be simultaneously effective, avoidable, and distinctive. Due to its tripartite

¹³ As Sarah Conly notes, when we see libertarian paternalism in this light it appears to suffer '...the worst of both worlds' (2013, p. 8) by '...neither having its cake nor eating it' (2013, p. 32).

nature, I label this tension at the heart of the nudge project the *Choice Architect's Trilemma*.

This tension differs from the conflicting motivations of libertarian paternalism. Identifying it is a tricky task, however, due to the under-defined nature of nudging. The boundaries between what counts as a nudge and what does not have long been a point of contention. Indeed, Sunstein's recent warnings about the trap of abstraction follow directly from the minimal classifying conditions that much of the debate over nudging operates with. Many phenomena predictably alter our behaviour without resorting to incentives or sanctions. Further definition is required in order to provide a satisfactory depth of analysis and guarantee the compatibility of our arguments. Therefore, we must first provide further detail on how nudges function before seeking to understand the basis of this trilemma. In this section, I propose one possible model of how nudges function. This discussion is partly speculative and I accept the attractiveness of alternative possible explanations. However, further detail is required to identify this troubling tension and the following explanation is one way of clearly illustrating it.

Consider the main insight that choice architects draw upon when designing policies—the simple fact that we cannot consciously submit every decision that we make to a reasonably informed cost–benefit analysis. Rather, we must follow our gut instinct, rely on emotional insights, and internalise a whole host of heuristics in order to get by.

Much of our behaviour is governed by automatic intuitive thought that exists somewhere between direct perception and full-blown reflective judgement. The balance between our intuitive and reflective systems of thought is largely regulated by: (a) the amount of *attention* required to access both systems, and (b) the amount of *effort* required for either system to function.¹⁴ Conscious reflective judgements demand greater attention and effort than automatic intuitive reactions, which is why we tend to rely on the latter to guide much of our everyday behaviour.¹⁵ However, our attention is limited and our intuitive reactions are often habitually determined. This makes them difficult to control or modify upon reflection.

These features determine the basic relationship between our two systems of thought: we must reflectively monitor and correct our intuitive reactions, yet the reflective judgements required for this monitoring and correction demand greater attention and effort than their intuitive siblings. Each of us instinctively understands this basic relationship of give and take between our systems of thought, and we tend to succeed in employing both systems advantageously. Although bad habits and mistakes in decision-making are common, each of us can recognise the prudence in prioritising our reflective efforts toward more important decisions where possible.

The finite nature of our cognitive capacities imposes this limitation on us by making our attention and effort prized resources that we cannot expend infinitely. Our decisions are subject to cognitive scarcity. We must decide (where possible)

¹⁴ For discussion of the diversity of attentional functions and a defence of its selective qualities, see Wu (2014).

¹⁵ This insight also explains why intuitive processes coexist more easily than reflective processes, which often disrupt each other (Kahneman 2003, p. 1451).

when best to employ these resources, which decisions to reflectively monitor and correct, and which not to. This trade-off imposes an opportunity cost on our decision-making. Conscious reflection on one choice leaves us less likely to be able to reflect on another choice because we have less effort to expend on it, less attention to pay to it, and less time to give it our consideration. Ignoring this trade-off and trying to do too much risks distraction, cognitive overload, and exhaustion.

This trade-off creates an important crowding-out effect in our psychology: when a decision requires too much attention, effort, or time to monitor and correct, then our reflective system cannot function as it otherwise would. Some reflective decisions require more attention and effort than we can give them due to our circumstances. When this happens, we are left to rely on our intuitive thought as our only effective guide. Because the opportunity cost in question results from the competition between different modes of thought for the same finite cognitive resources, let us refer to it as a *Friction Cost*.

Friction costs arise whenever our systems of thought compete for the same limited mental reserves rather than working together harmoniously to inform our decision-making. Every decision that we make is subject to possible friction costs and, as noted, we commonly negotiate these costs without significantly damaging our prospects. Many factors inform this cost, including time pressures, the number of concurrent reflective tasks, our experience and practice at the task in hand, and our mood (Kahneman 2003, p. 1467). Each one of these factors can divert our attention and make reflective judgements more effortful.

One possible way of understanding the distinctive character of nudging is to argue that choice architects *intentionally impose friction costs on decision-makers in order to shape their behaviour*. Choice architecture can improve our unreflective choices by nudging us toward certain beneficial outcomes by imposing friction costs on us to make both (a) certain intuitive outcomes, and (b) the process of reflective monitoring and self-correction itself more psychologically costly than they would otherwise be. For example, we incur friction costs when we seek to reflect upon our apparent eagerness to pick up the healthier lunch options in the cafeteria or to select the savings plan that appears most beneficial to us when we are considering our future financial prospects. When we lack the attention, effort, or inclination to reflect on our intuitive reactions (perhaps because we are in a rush to sate our hunger or know little about financial risks), then we are more likely to rely on a nudge and so policy-makers are better able to shape our behaviour without resorting to coercion.

Sunstein and Thaler's call to employ our everyday blunders and biases as intuitive levers of thought in policy-making is an acknowledgement of the nature and utility of friction costs. By imposing these costs in particular ways toward particular goals, policy-makers may reduce the likelihood that we reflectively monitor and self-correct our behaviour, and make it more likely that when we rely on our intuitive system of thought for guidance we benefit from doing so. As the latter system of thought is more easily (non-coercively) manipulated than the former, the inventive use of friction costs promises to allow policy-makers to change our behaviour for the better without coercively prohibiting particular actions. As we have seen, this is the basic principle of nudging. Thus, I suggest that

we can conceivably understand the distinctive nature of nudging through what I have named friction costs.

Establishing the Trilemma

The use of friction costs as a policy tool illustrates a second broader conceptual objection to nudging. Policy-makers can increase the effectiveness of their nudges by selectively imposing friction costs on our decision-making. However, greater friction costs make it more likely that we will automatically comply with a policy.¹⁶ This is because the larger the friction costs imposed by a policy, the more attention and effort we require to respond reflectively to its effect, and thus the less likely it is that we will be able to do so. Some nudges gain their effectiveness by constraining our reflective choices. However, by definition, nudges should not impose significant costs on our reflective decision-making because they must be cheap and easy to avoid. Choice architecture's fabled ability to change our behaviour in a relatively benign manner is famously central to its distinctive attractiveness.

Our initial study of the mechanisms of nudging highlights a tension between these two claims. Nudges guide our decisions by imposing friction costs on us. By doing so, nudges make it more likely that our systems of thought will compete rather than cooperate (due to cognitive scarcity). Nudges are distinctive because proponents claim that they are both effective and avoidable. However, effectiveness requires choice architects to influence our intuitive choices and avoidability requires decision-makers to have reflective choices. Yet one of these often comes at the cost of crowding out the other due to the friction costs involved. Choice architects will either shape our decisions to make reflective judgement less likely and intuitive thought more beneficial, or they will shape our decisions to make reflective judgement more likely and intuitive thought less beneficial. They cannot do both in each individual instance.

Something must give—the nature of friction costs makes it likely that nudging is less effective, less avoidable, or less distinctive than it at first appears. In what remains, I test this hypothesis. I distinguish between three types of nudge that employ friction costs in different ways. I then propose a problem for each type of nudge. If each problem proves accurate then we must conclude that any defence of nudging as a distinctive, effective, and avoidable form of policy-making is conceptually flawed and potentially oxymoronic.

¹⁶ This general relationship will not hold in all circumstances. First, if friction costs increase to the point of exhaustion then a nudge will fail if the person cannot decide at all. Second, large friction costs may draw attention to covert nudges, possibly diminishing their effectiveness. Much of the interplay between these effects will be contextually determined, which is why I restrict my analysis to the study of likelihoods rather than causal certainties.

Acute Nudges and the Problem of Avoidance

I have suggested that understanding how policy-makers employ friction costs to shape our behaviour is a plausible way to understand the distinctive character of nudging. In this section, I explore the problem facing nudges that impose friction costs in a way that crowds out our reflective reasoning. Returning to Thaler and Sunstein's initial definition, we see that nudges should shape our decision-making situation towards some predictable end without forbidding options or significantly changing our economic incentives. The resulting intervention must be easy and cheap to avoid (Thaler and Sunstein 2009, p. 6).¹⁷ One way of fulfilling this definition is to change our decision-making behaviour *solely* through shaping our automatic intuitive decisions. These interventions are maximally distinctive from traditional incentives and sanctions because they seek to change our behaviour through purely intuitive means. For this reason, let us label this sub-class of interventions *Acute Nudges*.

In order to shape our decisions via purely intuitive means, acute nudges must separate out our two systems of thought. Choice architects can achieve this by imposing friction costs that are large enough to guarantee that we lack the effort or attention required to reflectively judge between options in the context of our decisions. If this intervention is successful, it will fully crowd out our reflective judgements concerning the decision at hand and leave us to decide intuitively or not at all. Having crowded out our reflective reasoning, acute nudges then exploit our heuristics and biases to guide our intuitive behaviour towards a particular outcome.

Instances of choice architecture that impose significantly high friction costs on our reflective reasoning will qualify as acute nudges. These interventions are most likely to succeed in circumstances when reflective monitoring and correction is already difficult. Examples of such interventions include successful priming effects that use unconscious influences to make certain decisions more likely, default rules acting as safety nets in complex and time-sensitive decisions, and the intentional framing of emotive and unfamiliar information under circumstances of clear risk or danger (such as medical circumstances). In these cases, salient circumstantial factors already impose large psychological costs on our reflective reasoning. Acute nudges will likely exploit and exaggerate these background factors to bypass our reflective system, leaving decision-makers to rely on the nudge towards a specific outcome.

We should expect acute nudges to be relatively rare. This small group of interventions remains significant, however, because these interventions provide us with the clearest illustration of the tension between, on the one hand, distinctiveness and effectiveness, and on the other hand, avoidability. Acute nudges are distinctive because they shape our behaviour through purely intuitive means. To do this effectively, a nudge must crowd out the reflective monitoring and self-correction

¹⁷ See also: 'We do not have a clear definition of "easily avoided".... It would be arbitrary and a bit ridiculous to offer an inflexible rule to specify when costs are high enough to disqualify a policy as libertarian, but the precise question of degree is not really important. Let us simply say that we want those costs to be small' (Thaler and Sunstein 2009, p. 249).

required for decision-makers to wilfully avoid a nudge if they wish. We can name this concern the *Problem of Avoidance*.

The problem of avoidance is not the traditional concern over whether nudges are coercive or not. Acute nudges do not coercively remove options or threaten any significant financial costs. Indeed, these policies will not provide decision-makers with reasons at all because they seek to crowd out our reflective thought (rendering simultaneous incentives or sanctions self-defeating). We can escape the effects of acute nudges more easily than standard coercive interventions for this reason. For example, we will resist an acute nudge in cases where the intervention crowds out our reflective reasoning but our instinctive behaviour is too firmly established by our habits to be changed. In these cases, acute nudges will fail to change our behaviour in a way that coercion would have successfully changed. Here we can see the difference between acute nudges and coercive sanctions—friction costs prohibit the *reflective* pursuit of decision outcomes rather than simply removing that option from the table. Acute nudges make it more difficult to reflect on outcomes but not necessarily more difficult to pursue them.

This difference suggests that the problem of avoidance is narrower than the problem of coercion. Our worry stems from the fact that non-reflective avoidance arises by chance (e.g. when our intuitive behaviour is habitual rather than chosen on reflection). Acute nudges may be effective and distinctive. However, they prevent the reflective act of will required for a decision-maker to avoid the nudge consciously and wilfully. Therefore, acute nudges do not restrict our freedom of choice *simpliciter*. They do not prevent us from completing certain actions or pursuing particular opportunities. Indeed, in some lucky cases, we can avoid them. However, acute nudges do restrict our *freedom of reflective choice* and prevent us from reflectively deciding to act in these ways. Our reflective decision-making is constrained by the large friction costs imposed on us and we are unable to avoid the intervention wilfully.

According to the problem of avoidance, we can object that the significantly high friction costs imposed by acute nudges make it unreasonably difficult for decision-makers to reflectively monitor and correct their behaviour. The choice architecture in question makes it too psychologically costly for us to do so given our finite psychological resources. When policy-makers employ friction costs so large that they prevent us from reflecting in this manner, their interventions constrain an important part of our freedom of choice by rendering us unfree to engage in reflective reasoning.¹⁸

The problem of avoidance offers us an important insight. By conflating freedom of choice with the freedom to choose reflectively, choice architects suggest that effective nudges are avoidable because they do not coerce. Our study of acute nudges shows that an intervention can be avoidable without being *wilfully* avoidable (e.g. by only being avoidable through chance). This distinction explains why proponents cannot simply rely on the presence of opt-out opportunities as a way of

¹⁸ This raises a further concern that acute nudges threaten our personal autonomy by constraining our reflective acts of will. While it is true that many nudges are compatible with our autonomy (Wilkinson 2013; Blumenthal-Barby 2013; Hanna 2015; Mills 2015), so-called acute nudges are not for this very reason.

securing avoidability. Interventions can constrain our decision-making even when they provide us with an opportunity to opt-out. If that opportunity requires too much attention and effort for us to use, then we cannot make wilful use of that opt-out even when it is offered to us.¹⁹

While acute nudges are extreme cases, the fact that friction costs can be so large that they crowd out our reflective reasoning shows us how any nudge can constrain our reflective decision-making to some extent. The more a nudge leverages our automatic reasoning, the more difficult it is for us to avoid it wilfully because the leveraging makes reflective choice more psychologically costly. The resulting problem of avoidance establishes the first horn of the choice architect's trilemma: some nudges impose friction costs that are large enough to crowd out our reflective reasoning. The resulting nudges are effective and distinctive, but they are difficult to avoid as they undermine our freedom of reflective choice.

Dual-System Nudges and the Problem of Distinctiveness

The problem of avoidance suggests that it will be difficult to wilfully avoid nudges that seek to be both distinctive and effective. This problem afflicts a small number of interventions that possess a morally troubling character. For this reason, choice architects may happily dismiss acute nudges as failing to fulfil the definition of a nudge. According to that definition, nudges should not be costly in the way that acute nudges are. Choice architects can deny that nudges must crowd out our reflective reasoning and instead suggest that nudges can shape our behaviour by leveraging *both* intuitive and reflective reasoning in tandem. These interventions will impose smaller friction costs on decision-makers than acute nudges because they do not seek to crowd out reflective thought entirely. Rather, they have the aim of guiding our intuitive reactions towards a particular decision outcome and then prompting us to reflectively endorse these reactions to secure the outcome in question. This form of choice architecture will not bypass or quell our reflective capacities. Rather, it will harness both systems of thought together. Because these policies attempt to shape our decisions by leveraging both systems of thought, let us name these interventions *Dual-System Nudges*.

As these interventions impose lower psychological costs than acute nudges, we might expect dual-system nudges to be perhaps the most common type of nudge. For example, we can classify many of the common cases discussed in the literature in this way (such as better feedback and incentive schemes). Nearly all social nudges (such as the famous 'Don't Mess with Texas' slogan) can be classified in this way. Examples of dual-system nudges include interventions that make certain reasons relatively more appealing by framing an offer in a more attractive light or by clarifying a threat in an especially salient manner. Although these interventions do impose traditional costs on decision-makers, they still fulfil Thaler and

¹⁹ Riccardo Rebonato (2012, pp. 8–9) explains this problem by distinguishing between *nominal* and *real* freedom of choice. Rebonato argues that nudges only protect our nominal freedom of choice (with an opt-out condition) at the cost of our real freedom of choice (which is restricted by the success of the policy).

Sunstein's initial definition because these costs are low enough to avoid counting as coercion or significant economic costs. Instead, they can impose a range of *mild* psychological, social, or economic costs in order to shape our behaviour. So long as these policies impose friction costs to emphasise the intuitive features of the traditional threat or offer in question, then proponents may defend them as nudges.

We find an example of this strategy in Yashar Saghai's (2013) revisionary defence of nudging. Saghai seeks to secure a morally robust form of freedom of choice when designing choice architecture. According to Saghai (2013, p. 489), a nudge preserves our freedom of choice when: (1) it preserves our choice set, and (2) it is fully or substantially non-controlling. These conditions are distinctive and secure avoidability, but we might worry that Saghai's nudges lack the power to be effective. How could an intentionally non-controlling choice-preserving intervention guide our decision-making towards a beneficial outcome? To resolve this concern, Saghai defines effectiveness in terms of leveraging both systems of thought up to (but excluding) 'full-blown deliberation'.²⁰

Rather than exclusively leveraging our automatic reasoning to secure policy goals (like an acute nudge), Saghai's nudges leverage both our automatic and reflective reasoning in tandem. This revision dissolves the tension between effectiveness and avoidability that afflicts acute nudges by changing the terms of effectiveness. This strategy allows Saghai to defend what I have labelled dual-system nudging. This defence is realistic and attractive for a number of reasons. Chief among them is that dual-system nudges avoid the moral concerns facing acute nudges over freedom of reflective choice.²¹ However, assuming that dual-system nudges are effective at changing our behaviour, critics might worry that Saghai's revision erodes the distinctiveness of nudging. Let us call this the *Problem of Distinctiveness*.

According to this objection, nearly all sensible evidence-based policy-making that is sensitive to realistic human decision-making will seek to leverage both of our systems of thought together. This approach is not novel. Successful policy-makers know that both what they offer (or threaten) and how they make that offer (or threat) matter to the likelihood of success. Few seek to limit themselves to just one method of influence. They can design incentives that both significantly change the payoffs of our options and that frame this offer in a way that appears intuitively attractive. However, these incentives are not nudges in any distinctive sense. What makes choice architecture distinctive is how it imposes friction costs to achieve policy goals. The problem of distinctiveness suggests that dual-system nudges impose these costs in the same way that most other successful policies do.

²⁰ For example: 'In fact, all nudges rely on what I call "shallow cognitive processes". I will use the term to cover both nondeliberative and incompletely deliberative processes that share three properties: (1) they are fast; (2) the "cognitive miser" is inclined to rely on them because they consume few resources; (3) they yield responses that are not the result of full-blown deliberation (ie, the exploration of a broad, if not always exhaustive, range of hypotheses for solving a problem)' (Saghai 2013, p. 489).

²¹ Additionally, because dual-system nudges provide reasons in an intuitively compelling manner, proponents might hope that dual-system nudges are better equipped to produce sustainable long-term behaviour change than acute nudges.

While Sunstein and Thaler's arguments trade on the competition between automatic and reflective reasoning to derive a novel policy approach, Saghai's argument trades away this relationship in the hope of proving some behaviourally-informed policies to be both effective and avoidable. To the extent that he succeeds in his task, his arguments encounter a different concern that should be instructive to any choice architect: nudges that impose significantly high friction costs threaten our capacity for wilful avoidance while nudges that impose relatively lower friction costs threaten to lose their distinctive character.

Most policy interventions seek to be both intuitively appealing and attractive upon reflection. To distinguish themselves from other evidence-based policy approaches in the manner that Sunstein and Thaler intend, nudges must wield friction costs in a distinctive manner while protecting freedom of choice. If they cannot do this, then nudges risk becoming indistinguishable from other traditional policies that match how people actually make decisions. This problem of distinctiveness establishes the second horn of the choice architect's trilemma: some nudges impose friction costs that are small enough to leverage both reflective and intuitive systems of thought without one crowding out and dominating the other. The resulting policies are effective and wilfully avoidable, but the goal of making reasons intuitively appealing is not novel. As a result, dual-system nudging appears to be a label that is so broad as to be indistinct.

Dual-Goal Nudges and the Problem of Effectiveness

Choice architects can respond to the problem of distinctiveness in one of two ways. First, they may increase the friction costs imposed on decision-makers. This distinguishes nudging from traditional interventions at the cost of re-stoking the problem of avoidance. Second, they may change the way that they are imposing their mild friction costs. Policy-makers cannot reduce friction costs to zero and still claim to be nudging citizens toward better choice outcomes.²² However, they can employ the mild costs used by dual-system nudges to guide our two systems of thought in different directions rather than guiding our two systems of thought towards one outcome in tandem. Choice architects may design interventions that guide our intuitive reactions towards a desirable outcome while simultaneously prompting us to reflect on whether we endorse this outcome or would prefer to exit. These interventions will guarantee that decision-makers possess both an opportunity to opt out wilfully and the necessary means to make use of that opportunity. Given these divergent goals, let us name these interventions *Dual-Goal Nudges*.

Examples of dual-goal nudges include active choice prompts and explicit debiasing. These interventions use friction costs to separate out our two systems of thought and encourage one to reflect on the other. Rather than imposing these costs in order to dominate one system with another, dual-goal nudges trigger both systems

²² A response that rejected the use of friction costs entirely would be a move away from nudging towards more traditional sanctions and incentives.

of thought in order to simultaneously lead us toward some desired outcome and prompt us to reflect on whether we would prefer to avoid that outcome.

By using friction costs in this way, these policies distinctively emphasise the decision-maker's capacity for exit, and are thus more wilfully avoidable than acute nudges. Further, this dual-goal use of friction costs is more distinctive than dual-system nudges. However, the Janus-faced nature of dual-goal nudges raises questions over their effectiveness. If an intervention imposes mild friction costs in order to lead decision-makers both towards and away from policy outcomes, then it remains unclear whether that policy effectively guides the 'decision-makers' behaviour at all. These policies use the distinctive mechanisms of friction costs in order to secure avoidability. But in doing so, they appear to lose their effectiveness. Let us call this concern the *Problem of Effectiveness*.

Choice architects may respond to this problem by arguing that dual-goal nudges lead us towards policy outcomes in the most morally acceptable manner. Although engaging people's deliberative capacities can be psychologically costly to them, it is necessary in order to treat them as rational agents. The argument continues that this morally constrained form of effectiveness is the only one worth its name, and thus policy-makers should not desire to impose greater friction costs than these. While this response is certainly admirable, it ignores the quandary at hand. The mere use of friction costs is not the same as the effective use of friction costs, and the employment of friction costs to guide decision-makers towards two opposing outcomes is an ineffective use of those costs when the effects of both costs cancel each other out. Critics may reasonably suggest that an intuitive pull towards a policy goal and a simultaneous reflective push toward the exit offers no overall benefit to our decision-making. Although these policies use two different means of influencing us towards two different goals, it remains unclear whether either force brings us any closer to a decision. Indeed, many dual-goal nudges will simply increase the psychological costs of decision-making by bringing choosers closer to cognitive overload without bringing them closer to a beneficial outcome. For this reason, critics may dismiss dual-goal nudges as ineffective.

The problem of effectiveness establishes the third horn of the choice architect's trilemma: dual-goal nudges impose friction costs that are small enough to leverage both systems of reasoning without one crowding out the other. The resulting policies are distinctive and wilfully avoidable, but the aim of guiding decision-makers simultaneously towards and away from a policy goal risks ineffectiveness.

Conclusion

I have argued that two versions of the oxymoron objection continue to stalk the debate over the possibility and permissibility of nudging. I have suggested that the original motivational version of this objection, which denies the coherence of libertarian paternalism, still requires resolution. Further, I have identified a broader mechanical version of the objection that threatens nudging more generally. Many policy interventions satisfy Sunstein and Thaler's initial definition of a nudge. We can further explain the distinctive features of these policies according to how they

employ what I have called friction costs. This explanation seeks to show that although nudges do not rely on the traditional social or economic costs involved in incentives and sanctions, they do rely on certain psychological costs to shape our behaviour. Thinking of nudges in terms of friction costs helps to focus on this important feature.

This explanation allowed us to categorise nudges according to how they employ friction costs. I then suggested that each category of intervention faces its own problem. Nudges are supposed to be distinctive because they are both effective and avoidable, but closer study suggests this initially appealing definition rings false. Few nudge interventions will simultaneously enjoy all of these characteristics.

This is due to the existence of three related problems:

The Problem of Avoidability—Imposing significantly high friction costs offers policy-makers a distinctive method of effectively changing our behaviour. However, these costs crowd out our reflective system of thought and make it difficult to avoid policy interventions wilfully.

The Problem of Distinctiveness—Imposing moderate friction costs in order to trigger both systems of thought towards the same policy goal promises to effectively change our behaviour in a wilfully avoidable manner. However, these interventions are indistinguishable from other forms of evidence-based policy interventions.

The Problem of Effectiveness—Imposing moderate friction costs in order to trigger both systems of thought in different directions promises to secure wilful avoidance through distinctive means. However, these interventions are an ineffective means of securing policy goals.

These problems create:

The Choice Architect's Trilemma—A nudge will be either wilfully unavoidable, indistinct from other policies, or ineffective at securing policy outcomes.

This trilemma sets up the path between three pitfalls that choice architects must navigate. It also explains (and supports) Sunstein's recent call to move away from general abstract discussion of nudging. If nudges shared a more detailed series of definitional characteristics, then such a discussion would be useful. However, in the absence of a more detailed definition that avoids the conceptual objections raised here, general abstract discussion is difficult.

The choice architect's trilemma establishes a difficult choice for academics: either we need further definitional clarity to engage in abstract argument or we must acknowledge that the definition of a nudge is largely vacuous and we should proceed with a more specific case-by-case debate. The trilemma also establishes a difficult choice for policy-makers: if nudges are to be treated as a distinctive type of policy intervention, then they are either not the cheap and effective silver bullets that policy-makers desire (because they are less effective than advertised), or they

are more restrictive than commonly accepted (because they are more difficult to wilfully avoid than advertised).

So long as these tensions remain unresolved, critics can continue to dismiss the nudge project as oxymoronic because few interventions can claim to possess all three characteristics that nudges should possess. Most nudges will fall foul of one of the problems identified above. How problematic these problems turn out to be will depend on how we value these conflicting elements and the extent to which we wish to engage with the issues that they raise. However, in order to make this evaluative judgement, we must first understand the conceptual terrain that academics and policy-makers are attempting to traverse. I have argued that this terrain is far more treacherous than many fellow travellers realise.

Acknowledgements The author would like to thank Ben Colburn, Richard Moorhead, Fay Niker and the rest of the participants of the Behavioural Public Policy: Theory and Practice workshop held at UCL. Particular thanks are owed to Muireann Quigley whose concerns inspired much of this article.

Compliance with Ethical Standards

Conflict of interest The author declares that he has no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Arneson, R.J. 2015. Nudge and shove. *Social Theory and Practice* 41 (4): 668–691.
- Blumenthal-Barby, J.S. 2013. Choice architecture: A mechanism for improving decisions while preserving liberty? In *Paternalism: Theory and practice*, ed. Christian Coons, and Michael Weber, 178–196. Cambridge: Cambridge University Press.
- Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. Regulation for conservatives: Behavioural economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review* 151 (3): 1211–1254.
- Conly, S. 2013. *Against autonomy: Justifying coercive paternalism*. Cambridge: Cambridge University Press.
- Feinberg, Joel. 1986. *The moral limits of the criminal law, vol. 3., Harm to self*. New York: Oxford University Press.
- Grill, Kalle. 2014. Expanding the nudge: Designing choice contexts and choice contents. *Rationality, Markets and Morals* 5: 139–162.
- Guala, Francesco, and Luigi Mittone. 2015. A political justification of nudging. *Review of Philosophy and Psychology* 6 (3): 385–395.
- Hanna, Jason. 2015. Libertarian paternalism, manipulation, and the shaping of preferences. *Social Theory and Practice*. 41 (4): 618–643.
- Hayek, Friedrich August Von. 2006. *The constitution of liberty [1960]*. London: Routledge Classics.
- Kahneman, Daniel. 2003. Maps of bounded rationality: Psychology for behavioural economics. *The American Economic Review* 93 (5): 1449–1475.
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. London: Penguin.
- Kahneman, Daniel, and Amos Tversky. 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.

- Kapsner, Andreas, and Barbara Sandfuchs. 2015. Nudging as a threat to privacy. *Review of Philosophy and Psychology* 6 (3): 455–468.
- Kelly, Jamie. 2013. Libertarian paternalism, utilitarianism, and justice. In *Paternalism: Theory and practice*, ed. Christian Coons, and Michael Weber, 216–230. Cambridge: Cambridge University Press.
- Korobkin, Russell. 2009. Libertarian welfarism. *California Law Review* 97 (6): 1651–1685.
- Mill, J.S. 1991. *On liberty [1859]*. Oxford: Oxford University Press.
- Mills, Chris. 2015. The heteronomy of choice architecture. *Review of Philosophy and Psychology* 6 (3): 495–509.
- Mitchell, Gregory. 2005. Libertarian paternalism is an oxymoron. *Northwestern University Law Review* 99 (3): 1245–1278.
- Nozick, Robert. 1974. *Anarchy, state, and utopia*. Oxford: Blackwell.
- Oliver, Adam. 2013a. Introduction. In *Behavioural public policy*, ed. Adam Oliver, 1–15. Cambridge: Cambridge University Press.
- Oliver, Adam. 2013b. From nudging to budging: Using behavioural economics to inform public sector policy. *Journal of Social Policy* 42 (4): 685–700.
- Oliver, Adam. 2015. Nudging, shoving, and budging: Behavioural economic-informed policy. *Public Administration* 93 (3): 700–714.
- Rebonato, Riccardo. 2012. *Taking liberties: A critical examination of libertarian paternalism*. Hampshire: Palgrave Macmillan.
- Saghai, Yashar. 2013. Salvaging the concept of nudge. *Journal of Medical Ethics* 39: 487–493.
- Smith, Matthew, and Michael McPherson. 2009. Nudging for equality: Values in libertarian paternalism. *Administrative Law Review* 61 (2): 323–342.
- Stanovitch, Keith, and Richard West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences* 23: 645–726.
- Sunstein, Cass R. 2014. *Why nudge? The politics of libertarian paternalism*. Yale: Yale University Press.
- Sunstein, Cass R. 2015. Nudges, agency, and abstraction: A reply to critics. *Review of Philosophy and Psychology* 6 (3): 511–529.
- Sunstein, Cass R., and Richard H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review* 70 (4): 1159–1202.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.
- Wall, Steven. 2009. Self-ownership and paternalism. *Journal of Political Philosophy* 17 (4): 399–417.
- White, Mark D. 2013. *The manipulation of choice: Ethics and libertarian paternalism*. New York: Palgrave Macmillan.
- Wilkinson, Tim. 2013. Nudging and manipulation. *Political Studies* 61: 341–355.
- Wu, Wayne. 2014. *Attention*. New York: Routledge.