Are Health Nudges Coercive?*

Abstract

Governments and policy-makers have of late displayed renewed attention to behavioural research in an attempt to achieve a range of policy goals, including health promotion. In particular, approaches which could be labelled as ‘nudges’ have gained traction with policy-makers. A range of objections to nudging have been raised in the literature. These include claims that nudges undermine autonomy and liberty, may lead to a decrease in responsibility in decision-making, lack transparency, involve deception, and involve manipulation, potentially occasioning coercion. In this article I focus on claims of coercion, examining nudges within two of the main approaches to coercion—the pressure approach and the more recent enforcement approach. I argue that coercion entails an element of control over the behaviour of agents which is not plausibly displayed by the kinds of serious examples of nudges posited in the literature.

DOI: https://doi.org/10.1007/s40592-014-0008-5

1. Introduction

We are, according to behavioural psychologists and economists imperfect, decision-makers. This is because we are all prone to a range of cognitively-mediated effects which impact on our behaviour and decision-making. As such, the way in which we make choices and decisions is affected by a range of cognitive quirks or stumbles.¹ These stumbles can arise as a result either of our automatic and reflexive cognitive processes or our more conscious and reflective ones.² Changing the contexts in which we make decisions – the ‘choice architecture’ - can harness or eliminate our cognitive biases and heuristics (rules of thumb), thus altering the probability that we will behave in one way as opposed to another (Thaler and Sunstein 2008; Sunstein, 2014). Potential alterations to the choice architecture to influence behaviour may include strategies which encompass so-called ‘nudges’. The paradigm example of nudging presented in the literature involves the layout of cafeteria queues (Thaler and Sunstein 2008: 1-3). The manner in which food bars in the cafeterias are organised influences the choice of food which ends up on our plates. Research by Rozin and colleagues (2011), for example, demonstrates that food at a salad bar which is easier to access (closer to us) is chosen more frequently. Additionally, people take larger servings of food which is served with a spoon rather than tongs. Given this,

¹ I am grateful to Wouter Kalf, Wendy Rogers, and the editors for their comments on earlier drafts of this manuscript. Thank you also Stephen Wilkinson and other attendees at the Analytic Bioethics is Europe Workshop in Ghent May 2014. Albeit in a slightly different context I had conversations with them which helped with my thoughts here.

² I take the reference to ‘stumbles’ from Amir & Lobel 2008.

² Otherwise termed system 1 and system 2 processes.
the simple measure of changing the layout of the cafeteria (salad bar) will affect the food choices made. In this way the cafeteria choice architects can ‘nudge’ their patrons.

The potential to apply findings such as these to formulate public policy has received a great deal of attention from government and policy-makers. For example, the United Kingdom has had a Behavioural Insights Team (BIT) since 2010, the United States Social and Behavioural Sciences Team was set up early in 2014 (Steverman, 2014), and in Australia BIT is advising the New South Wales Department of Premier and Cabinet. There is also support at European level for running behavioural studies related to European Commission services (van Bavel et al. 2013). Yet, despite the enthusiasm for behaviourally-informed policy-making, attempts by government and regulators to alter the behaviour of citizen’s using nudge-type strategies have drawn opprobrium from certain quarters. A range of objections to nudging have been raised in the literature. These include claims that nudges undermine autonomy and liberty, may lead to a decrease in responsibility in decision-making, lack transparency, involve deception, and involve manipulation, potentially occasioning coercion (see generally Bovens 2009; Hausmann & Welch 2010; White 2011; Yeung 2012; and White, 2013). In this article I focus on claims of coercion. Consider the following comments from Mark D. White, whose recent book deals with some of the ethical issues regarding the use of choice architecture to influence behaviour. He contends that “because [nudges] are designed with the express purpose of manipulating people’s decision-making processes to change their behaviour in pursuit of interests that are not their own, I argue that nudges are coercive” (White 2013, 91). He claims that nudges are coercive where they “involve steering people towards making choices to promote ends and interests other than their own” (White, 2013, 93-94). Another commentator, Brett Caloia, argues that to the extent that nudges impose psychic costs (e.g. stress and other forms of “cognitive discomfort”) on the nudgee these “costs can have a profound enough effect on the agent’s life as to be coercive” (Caloia 2014, 23 & 18).

In this article I examine whether intentional alterations to the choice architecture (by policy-makers and other actors) can reasonably be said to amount to coercion. I conclude that concerns that nudges are coercive are overplayed. I do not discount the possibility that they could be, but this is not a feature necessitated by the types of (policy) interventions currently being discussed in the literature. To facilitate this examination, let us look

4 The UK team, in particular, has produced a number of reports across a range of areas, including an early one on the application of behavioural insights to health. This contained a range of potential policy applications of behavioural research, including in the areas of smoking, alcohol consumption, organ donation, and diet, weight loss, and exercise (BIT 2010, BIT 2013).
5 Emphasis removed.
first at some possible scenarios in order to consider the kinds of cognitive biases and other effects which might influence our health-related decisions, choices, and behaviours.

2. Influencing persons through choice architecture

As already noted in the introduction the kinds of strategies at issue are those which attempt to influence behaviour by altering the contexts in which we make decisions. Interventions which harness, mitigate, or eliminate these are broadly captured by the concept of nudging. Consider, for example, Jane who has been diagnosed with gallstones. She has a consultation with her surgeon who explains the risks to her. The surgeon might tell her that there is a 97% probability that the surgery will go well and everything will be fine. On the other hand she could tell Jane that there is a 3% chance of complications arising from the surgery. The statistical risk is objectively the same, yet the way in which the surgeon frames the statistical information regarding the risks may influence whether or not Jane consents to the surgery. This illustrates the framing effect. The way information is presented to us changes the probability that, out of a set of choices, we will choose in one direction rather than another (Tversky & Kahneman 1981; Marteau 1989; Almashat et al. 2008). Jane’s decision about whether or not to go ahead with the surgery might also be affected by whether or not she can easily call to mind any examples of similar surgery. So if her friend Peter has recently had surgery, this may affect her assessment of the probability of complications or success consequent on the surgery (regardless of the actual risks involved). This is known as the availability heuristic (Tversky & Kahneman 1974, 1127-1128).

Status quo bias, another potentially pertinent cognitive effect in the health context, captures the observation that individuals are in general unlikely to make active changes away from current default options (Kahneman et al 1991). As such, it seems (theoretically at least) to support the contention that opt-out systems of organ donation could yield higher numbers of organs for transplantation.8 Think about Timmy who has lived in England for most of his life. Despite regularly seeing appeals for people to join the organ donor register he never has. He has in fact never really given it much thought. Due to a change of job he moves to Wales, where a law has recently been enacted introducing an opt-out system of donation.9 Under this system the default is that residents in Wales are opted-in to be organ donors after death, unless they actively opt out by putting their names on an opt-out register. Timmy never puts his name on this register.

8 Whether or not opt-out systems actually have a significant impact on the numbers of organs available for transplantation is a complex question. Many factors impact on this, even if more potential donors are available. See [citation removed for anonymity] and [citation removed for anonymity].
Behavioural research has shown that social norms are powerful. The fact that we are influenced by what others do, could be used, for example, to try and decrease alcohol intake in particular social groups. Consider Clarissa who studies physics at Monash University. She regularly goes on nights out with her friends and more often than not drinks in excess of the weekly recommended average. She thinks that this is what most students do. A new poster campaign regarding alcohol consumption has started on the University campus. The posters point out the average consumption of alcohol by university students. This is less than Clarissa’s average intake and encourages her to reduce her own consumption. This change in behaviour could be explained not only by the power of social norms (that is, knowledge of what her peers do), but also by a process of anchoring and adjustment (Tversky & Kahneman 1974, 1128-1130). Reference points influence us. As such, when presented with a new (lower) incidence of alcohol intake amongst students, Clarissa may anchor to this new information, adjusting her reference point and drinking downwards.

The overarching lesson from behavioural research which is illustrated in these scenarios is that “contextual influences” (be they informational, environmental, experiential, and so on) can affect the decisions we make (Sunstein & Thaler, 2003, 1161). In each scenario we can say that our protagonists have been nudged, that some aspect of the choice architecture influenced their decisions and behaviours. They trigger what Yashar Saghai calls “shallow cognitive processes” (Saghai 2013, 489). By this he means that “(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” (Saghai 2013, 489). In this respect they largely fly under the cognitive radar, often (although not always) bypassing our rational processes. We can see this, for example, in the scenario with Timmy. He has not been rationally persuaded that he ought to become an organ donor; he has not arrived at this position based on reasons offered to him either by himself or others. Instead the change in the legislative default which comes with his move to Wales takes advantage of the fact that “defaults are sticky” (Amir & Lobel 2008, 2120). Timmy has not been persuaded to donate by force of reason; he has not been persuaded at all. The opt-out default works on the basis of his inertia. Of course, the mere presence of a non-rational influence does not preclude other reason-based influences. Perhaps, after all, Timmy had been paying attention to the various educational campaigns over the years and is rationally persuaded that he ought to put his name on the organ donor register. He has just never gotten around to it. As a happy coincidence his move to Wales has operationalised the conclusion he already reached.

If we take the findings from various behavioural studies seriously, we are unlikely in these scenarios to say that our protagonists have not been influenced by the contextual elements which surround them. Questions arise, however, about what this influence amounts to and whether nudges are ethically acceptable forms of influence. There is not space in this article to examine all relevant aspects which might bear on these broad questions. However, what I do is to take seriously the claim, exemplified by White’s
comments in the introduction, that nudges are coercive. Recall that he says that the coercive effect of nudges lies in the fact that they “involve steering people towards making choices to promote ends and interests other than their own” (White 2013, 93-94). It would I think be plausible to say that the characters in my scenarios could have been ‘steered’ towards ends which may not be their own. In the surgery scenario, for example, perhaps it turns out that Jane is a partner in a firm of clinical negligence lawyers and asks the surgeon some astute questions about the risks involved in the surgery. The surgeon duly explains the risks to her, but, acting defensively, frames them negatively. If Jane subsequently refuses to consent to the surgery, we might think that she has been steered towards the surgeon’s ends (protecting against litigation) rather than her own (relief from her symptoms). Likewise, we might think that Timmy and Clarissa had their decisions nudged in the direction of the choice architects’ choosing – more organs and less alcohol consumption.

Nevertheless, coercion is not the same as ‘steering’ people or their choices, regardless of who is doing the steering or the ends towards which they are steered. As I will demonstrate in what follows, coercion is a much more pernicious type of influence and it does a disservice to the wrongness of coercion-proper to label nudges as coercive. In so doing I will not attempt to lay out or justify a thorough-going account of coercion or coercive action. There is not space here for such an endeavour and, in any case, there is already a large and comprehensive literature debating numerous aspects of such accounts. Instead I aim to set out in very general terms some of the features of standard cases of coercion and examine how well some examples of nudges track these. While I accept that the nuance and detail of these features is subject to (often extensive) debate, my claim is that the types of interventions or policies at issue are far removed from standard cases. As such, we should be hesitant to say that they count as coercion.

3. Considering coercion

3.1 Defining nudging?

One response to claims like White’s is a stipulative one. We could say that nudges by their very definition are not coercive. Indeed Richard Thaler and Cass Sunstein’s original proposition has this implication. They specifically define a nudge as “an aspect of choice architecture that alters people’s behaviour in a predictable way without forbidding any options” (Thaler & Sunstein 2009, 6). Meanwhile Saghai says that “A nudges B when A makes it more likely that B will ψ, primarily by triggering B’s shallow cognitive processes, while A’s influence preserves B’s choice-set and is substantially non-controlling (ie, preserves B’s freedom of choice)” (Saghai 2013, 491). These characterisations can be taken as having both descriptive and

normative elements. They are descriptive in the sense that they can be taken as simply describing the mode of action of nudges. Nudges are those, and only those, influences that leave our choice set open and do so via a non-controlling method. They are, however, moralised to the extent that they entail the normative moral criteria upon which they are purported to be justified. For Thaler and Sunstein at least, nudges are defended (partially) on the basis that they do not infringe the liberty of those who are nudged. Now if we think that one of the hallmarks of coercion is the foreclosure of options to the coercee or that they infringe individual liberty in some way, then this seems to automatically rule out nudges when defined in relation to the available choice-set. Likewise, if we think that the relevant feature of coercive action is that it is controlling from the perspective of the coercee, then defining a nudge as substantially non-controlling gets us around the issue. Taking this approach there is no need for us to ask if nudges are coercive, we know they are not. Instead the question becomes whether the types of influences and interventions being proposed are nudges narrowly-conceived or are they something else? I do not take this stipulative route here. Let me explain why.

In debates regarding nudges there is often terminological slippage. Sometimes the term ‘nudge’ is used when the commentator simply means influence, or at the very least some other sort of influence not covered by Thaler and Sunstein’s original characterisation or Saghai’s refined one (both given above). Given this tendency definitional precision is surely necessary, especially if commentators are to avoid talking past one another. Nevertheless, concerns regarding the behavioural interventions at issue (especially charges of coercion) cannot be taken seriously if I use a definitional sleight of hand early on. Therefore, I adopt a broad (perhaps overly so) idea of nudging for the purposes of the analysis in this article. I essentially proceed by example and use ‘nudge’ as an umbrella term for any of a range of approaches which intentionally use choice architecture to influence behaviour. Such a route is perhaps also more faithful to the position taken by those who criticise nudging since they are often interested in an array of interventions which draw on behavioural research, rather than those which meet strict definitional criteria.

We saw in the last section that while nudges and rational persuasion are not mutually exclusive, they operate differently. Therefore, although I cast the nudge net quite broadly in order to consider possible instances of coercion, I do exclude rational persuasion. I also omit outright bans and other forms of choice removal due to exogenous conditions. For the purposes of this article I am agnostic on the question of whether or not these would count as coercion. It is not that they could not count as such, but that choice removal is not captured by the types of behavioural

---

11 See, for example, both Scoccia and Hausman & Welch who seem to do this (Scoccia 2014, 2 and Hausman & Welch 2010, 124, fn 4)

12 In the case of the law, for example, Hughes notes the long-standing divergence of views between legal and political philosophers. The former tend to deny that law is coercive, while the latter group often readily accepts that it is (Hughes 2013, 231).
interventions at issue. I, therefore, do not include it in the general discussion of coercion that follows. Let us turn then to consider in the next two sections whether putative instances of nudging track some (or any) of the features of standard cases of coercion.

3.2 What does coercion look like?

Coercion is typically said to involve the intentional use of force or threats to compel a person to act and is often (although not universally) differentiated from instances of physical compulsion. Consider Timmy from the status quo bias and organ donation scenario. If when he moves to Wales, his next door neighbour calls around, holds a gun to his head, and says ‘your life or your kidney’, we might rightly say he has been coerced. This is a variant of the paradigm case of coercion (‘your life or your money’) often presented in the literature and exhibits some of the key facets of coercive action. Principally, it encompass circumstances where the coercer puts pressure on the will of the coercee in order to bring about some or other state of affairs (the ‘pressure approach’ to coercion). The coercer’s actions or threats are such that they force the coercee to do what is being requested; they are controlling from the perspective of the person being coerced. If we think that an example which involves threats to life look too extreme, we can amend the thought experiment so that the threat involves some other unpleasant option which Timmy would prefer to avoid. Imagine that his neighbour is in need of a kidney urgently. Instead holding a gun to his head and saying ‘your life or your kidney’, he threatens to evict him from his house. As it happens Timmy’s neighbour holds the mortgage for his house. He says that if Timmy does not agree to ‘donate’ a kidney, he will foreclose on it. Timmy is broke, cannot get a mortgage from a bank, and has barely enough money to feed his family. He is desperate and believes that other options (i.e. being homeless) look worse. He thus agrees to give his kidney to his neighbour. Even though it is not life-threatening, the threat made by the neighbour in this situation is controlling; Timmy will do what he is asked because the threat means that the consequences of refusal are dire. Timmy could choose to be evicted from his house, but as Joel Feinberg has noted, in such cases “one of the alternatives is so unreasonable that it is as if “he has no choice” but to opt for the other one” (Feinberg 1986, 151). In these scenarios Timmy prefers to give up his kidney rather than risk his life or being left without his home. While we could say that the threat provides him with a powerful reason to do as he is

---

13 Joel Feinberg, for example, draws such a line, arguing that compulsion-proper operates via direct force and leaves the person with no choice whatsoever, while coercion occurs when they are forced to choose the least-worst option (or to choose the ‘lesser of two evils’). See Feinberg 1986, 151 & 190-192. Whether such a distinction holds need not trouble us here since the non-rational influences under consideration unequivocally do not involve the use of direct physical force.

14 I follow Scott Anderson in his use of this term for these general approaches to coercion. See Anderson 2010. See also Lammond 2000, 44.

15 Emphasis added. As Scott Anderson puts it “for the coercee to choose otherwise would be contrary to practical reason.” Anderson 2011, 30.
asked, he has clearly not been rationally persuaded based on the merit and quality of arguments regarding the plight of his next door neighbour. Even where the threat made is an empty one, the situation can be appropriately viewed as coercive. All that is required is that the threat seems _credible_ to Timmy (Anderson 2011, 15 & Nelson et al. 2011, 7).

We might think that such cases are not a fair representation of what coercion might look like in the context of influences which trigger shallow cognitive processes. The sorts of circumstances that fall under the umbrella of nudge do not consist of gun-touting or mortgage-holding neighbours who want to get their hands on your kidney. This is something accepted by critics of nudge when they claim that nudges are coercive. White, for instance, says “[t]o be sure, nudges are not comparable to a mugger holding a gun to your head or a police officer standing ready to arrest you” (White 2013, 91). However, even less serious cases of coercion display a similar structure and features, even if they do not do so exactly or perfectly. To illustrate let us take a more trivial example than the ones just considered. In this scenario Timmy’s neighbour threatens to destroy his award-winning flower display, if he does not put his name on the organ donor register. The point of the threat is to get Timmy to do something he would prefer not to do, put his name on the register. If he has reason to think his neighbour will carry through on his threat, then he is not left with much choice but to comply if he wants to save his flowers. Even if Timmy would think, on reflection, that he ought to (or wants to) take this course of action, we might think that the means of getting him to do it is coercive; that is, the use of a threat. Thus, the markers of coercion appear to be present, albeit the consequences of non-compliance are not as severe as in the earlier cases. Other examples commonly supposed in the literature to be cases of coercion include:

... a drug dealer’s threat to withhold a regular customer’s supply unless he beats up another person; a threat to destroy someone’s beloved car unless he commits a murder; a dry cleaner’s threat to refuse to return a customer’s laundry unless she pays an unjust $10 premium; a blackmailer’s threat to reveal one’s infidelity unless one pays a substantial sum; and a millionaire’s offer to pay for surgery for a gravely ill child if only the child’s mother agrees to become the millionaire’s mistress (Anderson 2010, 3-4).

Although these clearly differ in their contexts and vary in the degree or severity of the coercion present, they all share the features highlighted by my earlier scenarios: (1) they involves the use of force or threat to put pressure on the will of the coercee; (2) the force or threat must be deemed to be credible by the coercee (they must believe that the coercer will follow through and that the threatened consequences will come to pass); and (3) the coercee is compelled to do (or not do) something which they would prefer not to - to ‘choose’ the ‘lesser of two evils’ as Feinberg puts it (although presumably they could be coerced into doing something which they would prefer to do all things considered, but just not do right now or not do if not being forced into it) (Feinberg 1986, 151). From these we can also note a fourth feature. Coercion is not simply (or just) a feature of an...
individual’s background circumstances, it involves the actions of an intentional agent (Scoccia 2014, 1). Certainly the circumstances in which a person finds themselves can contribute to our assessment of whether or not something amounts to coercion. However, absent some agent (or private or public agency/institution) exploiting such conditions, we tend to think of them as tragic or unfortunate rather than coercive per se. Finally, we should note that coercion is commonly said to entail a success condition. This comes from Robert Nozick’s exposition which most variations of the pressure account stem from. On this view coercion has only taken place if the pressure applied by the agent is successful in getting the coercee to do (or not) what is being demanded of them and can be said to have played a causal role in their so doing (Nozick 1969, Anderson 2014, 17-18). Of course, lack of success does not imply that there has been no attempt to coerce.

How then do these features sit with attempts to influence behaviour using techniques which harness or eliminate cognitive biases and other effects?

3.3 Nudges as coercive pressure?
In order to draw out some of the features ordinarily associated with coercion, I purposefully offered thought experiments at the extreme end of the scale. By starting with unambiguous paradigm cases of coercion we can see whether more moderate scenarios involving nudges track some (or any) of the elements of coercive action. What I suggest in this section is that, to the extent that we could claim that they do this at all, they do not do so in any thorough-going manner. Nudges track neither the general structure nor any individual features in a way that leaves us with a robust version of what coercion looks like.

First let us concede that there is at least one way in which nudging maps onto aspects relevant to an assessment of coercion: the involvement of parties who attempt to bring about particular outcomes. We could conceive of all manner of random (non-designed) aspects of the choice architecture as nudging us in one direction or another. However, in the specific context of policy and other planned interventions, we can grant that

16 Feinberg, for example, distinguishes between an agent making “an intervention on the stage rather than [being] part of the stage setting” (Feinberg 1989, 149).
17 Nozick’s account contained six elements which I reproduce in their summary from as set out by Anderson: “(1) P aims to keep Q from choosing to perform action A; (2) P communicates a claim to Q; (3) P’s claim indicates that if Q performs A, then P will bring about some consequence that would make Q’s A-ing less desirable to Q than Q’s not A-ing; (4) P’s claim is credible to Q; (5) Q does not do A; (6) Part of Q’s reason for not doing A is to lessen the likelihood that P will bring about the consequence announced in (3)” (Anderson 2011, 15). This was intended to be a non-moralised account and stands in contrast to later accounts such as Alan Wertheimer’s who argued that coercion needs to be measured against a moral baseline. See generally Wertheimer 1990, 206-221.
18 For a discussion on some aspects of the moral relevance of random versus designed choice architecture see Quigley 2013, 610-617.
there will be an agent doing the nudging (or designing the setting in which the nudging takes place). What is more, they will often be doing this with the explicit intention of getting the nudgee to take (or not) some or other course of action or decision. Here we need not view the relevant form of agency as arising purely at an interpersonal level (like that between Timmy and his neighbour). It can encompass relations involving private organisations or public agencies. It is, in certain contexts, reasonable to see them as displaying the requisite agency and intentionality for the purposes of coercion. Indeed, we often conceive of them thus (the relationship between state authority and coercion, for example, is subject to wide debate in the legal and political philosophy literature). Hence, both public and private actors could intentionally implement nudge-type interventions. Marketeers and government could, for instance, employ techniques which draw on the power of social norms: one to sell their products, the other to try to reduce alcohol consumption or smoking, and so on. It is also true that these techniques could be used to get individuals to take decisions or courses of action which, all things considered, they might prefer not to. Timmy in our status quo bias scenario, might in fact prefer not to be opted in to the organ donor register. This, however, is where the parallels appear to stop. Many activities by individuals or organisations involve them steering us in directions of their own choosing, but this alone does not amount to coercion.

The most obvious point of divergence is that nudges do not involve the use of force or threat to achieve their ends. Even if the purpose of some nudges is to cause us to choose, decide, or act in a particular way, it would seem to stretch our ordinary understanding of ‘force’ to apply it in these contexts. Think back to Jane and her surgical consultation from section two of this article. Regardless of whether or not the surgeon frames the risk of surgery positively or negatively, no force or threat will have been deployed. This is also true of Clarissa in the social norms and alcohol scenario and Timmy in the status quo bias example. Such attempts to influence behaviour are unlikely to amount to threat or force in even a minimal sense. By their very nature the mode of action of these behavioural influences is different. As noted earlier, they often fly under the cognitive radar. Threats and other methods of forcing are explicit propositions; something is being demanded of the coercee. Moreover, it is being demanded in a manner which makes it exceptionally difficult to resist. Imagine that a new government policy mandates that the cafeterias in all public institutions and buildings are to be laid out so as to nudge patrons to choose the healthy food. Behavioural research tells us that the probability that the cafeteria patrons will choose the healthy options has been altered, but we cannot reasonably think that they are being forced to choose. The choice of food in the queue is not an ‘or else’ proposition. The new food policy does not say ‘choose the healthy option or else you do not eat at all’. If they enter the cafeteria with the express intent of having their usual burger and chips, they may well still choose this option. They can easily resist the influence of the changed cafeteria layout. This supposition about resistibility seems to be reasonable, even taking into account the fact that for psychological and
other reasons some individuals may more easily be able to do so than others.¹⁹

Conceivably by focusing on force or threat we end up with a conception of coercion which is too demanding and which leaves us with too narrow a class of activities which could count as coercive. The mechanism of action by which either force or threat operate is through the exertion of pressure on the will of the coercee. But there are other ways in which this can be achieved. Opponents of nudges might feasibly claim that the effect of interventions utilising non-rational cognitive influences is to apply just such pressure to the will of nudgees. Indeed Caloia argues that certain nudges operate by imposing psychic costs on the nudgee; that is, while there is no force or threat as such, they cause the nudgee to alter their behaviour by causing them cognitive discomfort (Caloia 2014, 23). Examples of this given by him are the use of graphic images on cigarette package and the use of information about societal norms (Caloia 2014, 18-19). If we allow for an expanded understanding of pressure on an individual’s will, then we can re-configure our conception of coercion to also be wider in this respect. At least at first glance it looks like the pressure approach would not rule this out. According to Scott Anderson pressure approaches to coercion are “fairly ecumenical about what sorts of threats can be the source of such pressure; anything disadvantageous to the coercee’s interests seems fit to be used to cause a coercee to alter her actions” (Anderson 2010, 9). If Timmy did not care for his flower display, then there would be little point in his neighbour threatening to destroy them. It is only because doing this is disadvantageous to Timmy’s interests that any pressure is exerted by the threat. If in fact Timmy had decided earlier that he was going to get rid of his flowers and dig up his flowerbed, there would be no disadvantage to his interests (either immediate or longer term). Therefore, the requisite pressure would not be present. Taking this approach we could construct coercion along the lines of disadvantage to interests in non-threat scenarios. Doing this might capture the kinds of alterations to the choice architecture at issue when commentators criticise nudging.

We need to be clear, however, what disadvantage means in the context of coercion. One might point out that by pinpointing disadvantage we rule out the possibility of advantageous coercion, something which is at least theoretically possible.²⁰ This can be illustrated if we think about coercive offers. Consider that, instead of waving a gun at his head, Timmy’s neighbour offers to give him $50,000 once the kidney retrieval operation is over. If Timmy felt he had any other choice, he would not part with his kidney, but he does not so he does. He duly receives the cash after

---

¹⁹ Plausibly, regardless of the layout of the cafeteria, it may be even be harder to resist the allure of the burger and chips than the salad bar. For a discussion of resistibility in the context of his definition of nudges see Saghai 2013. As alluded to earlier, for him, nudges-proper must be easily resitible.

²⁰ Thank you to Michael Selgelid for this point and for prompting me to clarify my thoughts here.
the operation. This is enough to pay off his debts and for him and his family to live comfortably for a while. Because the influences which had a bearing on his decision were, for all intents and purposes, fully controlling, some might consider this to be coercive (e.g. Zimmerman, 1981 & Feinberg 1986, 229-268). Others would disagree and claim that such offers are perhaps irresistible, but do not amount to coercion. Alan Wertheimer and Franklin G. Miller, for example, say that “genuine offers do not coerce” (Wertheimer & Miller 2008, 389; see also Day 1992). I am not going to make an argument on this either way, but the purpose of the example is to show that, if we do think offers can count as coercion, then we are committed to the possibility of coercive action which can be advantageous to the coercee. There may even be classic threat-type situations which, although coercive, are advantageous. Clarissa’s parents, for example, might threaten that they will throw her out of the house and stop paying her university tuition fees if her grades do not improve (or she does not reduce her drinking). If the threat is effective we might well think that this is to her benefit, but that she had nonetheless been coerced. These examples are relevant to Anderson’s comment regarding disadvantage because, we need to differentiate between disadvantage all things considered and disadvantage at the point of the intervention. While the effect of the coercion might be a global benefit or advantage to the coercee, the coercion is only effective because of a more instant disadvantage at the point of intervention. It is this instant disadvantage that is causally linked to the coercive act. In effect it is part of the mechanism of action of the coercion. The coercer takes advantage of a disadvantage so to speak.

Under this disadvantage construction of coercion there is the potential for circumstances beyond the usual overt force or threat to be viewed as putting the relevant cognitive pressure on persons. It is, however, far from clear whether the kinds of hypothetical nudges described in the literature, or indeed any serious policy proposals, would count. Almost certainly the paradigm example of the cafeteria would not. When the cafeteria patron chooses salad rather than their usual burger and chips, there is no disadvantage being used as leverage to compel this choice, even if we construe the altered layout as constituting some sort of psychic pressure. Moreover, even if we were to accept that the choice architecture of the cafeteria is a form of pressure and accepted that it constituted a direct causative link, this is not enough. The pressure would need to be sufficiently controlling to warrant consideration as coercion. As such, this comes back to the issue of resistibility; coercive pressure is not easily resistible. The cognitive pressure exerted on the cafeteria patron does not seem to measure up in this respect. Nudge or no nudge, the burger and chips still look like a viable option here, something which does not appear to be so in standard cases of coercion.

It is, of course, possible that other nudge-type interventions do hit the mark in the terms just outlined. Situations which, as Caloia argues, impose cognitive discomfort on individuals look like prime candidates. His particular examples, as we have already seen are the graphic labelling of cigarette packages and strategies which utilise social norms. He says
“[f]orced confrontation with the consequences of smoking or overeating will tend to create negative emotions in agents” (Caloia 2014, 26). Further, he says that their coerciveness “ought to be judged by how confrontational the messages are, how difficult the feelings produced are to resist and how much of a negative impact these feelings impose on the life of the agent” (Caloia 2014, 26). Within the disadvantage-pressure thesis we could posit that the psychic discomfort produced from such interventions is both a pressure on the will of the agent and simultaneously constitutes the very disadvantage by which the pressure works to coerce. It is the discomfort itself which leads to the behaviour change. If this is correct then, theoretically at least, we might think that they are coercive. While this sounds plausible, there are two points that we need to bear in mind. First, if such strategies work in this manner, then they will only work for those in whom overt cognitive discomfort is created. Accordingly, they will coerce some and not others. This in itself is not a problem for pressure approaches, since it is merely trivially true that different people will respond in differing ways to pressure, be it force, threats, or otherwise. The second point, however, is again about resistibility. Even examples like the graphic cigarette packages seem unlikely to produce enough pressure for us to say that the person was compelled not to smoke. Much like the burger and chips within the altered choice architecture of the cafeteria, the consciously committed smoker would still be able to succumb to the allure of the cigarettes, despite the unsavoury packaging. It may be that there are other nudges that would fail the resistibility test, but the types suggested here would likely not.

3.4 Coercion, power, & being coercive

Conceivably the pressure approach to coercion is not particularly suited to depicting coercion in the context of behaviourally-informed public policy. For that reason, perhaps I could be accused of drawing on an account of coercion which is too strong or too exacting for capturing coercion in relation to nudging. Anderson has recently proposed an alternative understanding of coercion. This view, which he terms the enforcement approach, has as its focus the power disparity between agents: it “regards coercion as a kind of activity by a powerful agent who creates and then utilizes a significant disparity in power over another in order to constrain or alter the latter’s possibilities for action” (Anderson 2010, 1). He thinks that this conception has advantages over the pressure approach. He highlights how the pressure approach cannot account for much of the “social and political significance attributed to coercion” (Anderson 2010, 2). It cannot, for example, account for the fact that the power relations between parties are not static and can be impacted by a variety of factors, whereas the enforcement approach can. On his account the coercer “is either responsible for creating a power differential over the coercee, or else leverages some pre-existing differential in power” (Anderson 2010, 7). This encompasses modes of coercing characteristically associated with pressure approaches, such as force, violence, or threats, but also captures a broader range of powers. Power on this approach could be either brute physical power or that
conferred by being an office-holder or part of particular institutions (Anderson 2010, 12). The emphasis on power differentials covers situations where the “agent feels no overt pressure to do as the coercer demands” (Anderson 2010, 9). Hence this approach seemingly explains why certain laws could be seen as coercive, even though we feel no particular pressure from them or would follow their stipulations regardless; (i.e. perhaps due to a feeling of moral obligation) (Anderson 2010, 9). Anderson’s observations regarding power differentials are important in the context of the current discussion since, for critics of nudge, much of the ethical objections centre on attempts by government to apply the findings of behavioural research to the realm of public policy. If, as Anderson suggests, coercion involves the exploitation of a power imbalance by one agent over another in order to prevent or bring about some course of action, then it is at least plausible that it is pertinent to nudging in a way that the pressure account is not.

There are two potential ways in which this approach looks like it could capture coercion via nudge-type techniques or interventions. First, even if interventions which by-pass our reflective cognitive processes can be construed as pressure on our will (as we supposed in the previous section), they do not cause us to feel any overt pressure. Therefore, if claims of coercion are to have any bite in the context of nudging, we need an interpretation of coercion where this is not a requirement. This approach offers that. Secondly, both government institutions and private organisations can leverage power over individuals when they attempt to nudge them. This power can be viewed in relative terms; for example, a company may have a large marketing budget and experts in marketing who can utilise their knowledge of human behaviour in constructing their campaigns. There is a power imbalance between the company and individuals both in monetary and informational terms. Likewise, the machinery of government means that the state can wield power over its citizens in a variety of ways. Yet there is more to it than simply creating or leveraging pre-existing power imbalances. While the overt pressure might be not be an explicit requirement on the enforcement approach, there is still a focus on control. Coercion requires that a substantial element of control be exerted over the behaviour of the coercee. This is the implication of the emphasis on the relative power between actors. According to Anderson, “coercion generally disables (or threatens to disable) its target from being able to take effective countermeasures, or renders him unlikely to succeed or dangerously imprudent” (Anderson 2010, 28). In the nudge scenarios considered in this article the harnessing of particular cognitive biases has not rendered our protagonists unable to take countermeasures. Neither would it be foolish of them to take alternative courses of action. In each of the scenarios, although the probability that they would choose in one direction rather than another was changed, they could still have chosen differently. In the social norms scenario, for instance, Clarissa could still drink more. This option has not been removed (unlike, for instance, if there was a ban on alcohol on campus). While it might be imprudent for her to do so in health terms, it is not unwise in the sense that she might fear repercussions from the putative coercer for not reducing her intake (i.e. the government or university).
At this point two interrelated rejoinders to my arguments present themselves. First, it is more than likely that some behavioural techniques are in fact controlling. This could be the case either in the sense that their influence is not easily countered or because they have the effect that the nudgee believes that options have been foreclosed (even if in actual fact they have not been). I am open to the possibility that there are interventions which are being labelled as nudges and which fall into this category. If this is the case then we might rightly want to level the charge of coercion at them. Secondly, we might rescue charges of coercion by distinguishing between coercion and being coercive. It could reasonably be argued that critics are not claiming that nudges entail coercion per se, merely that nudges are coercive. Grant Lammond argues that to claim that something is coercive is to make a lesser charge than full-blown coercion. Here there need not be intent to coerce, but the effect of particular interventions may be coercive. Alternatively the requisite intent may be present, but the attempt to coerce might not be successful (Lammond 2000, 52 & Lammond 2013, 1-2). This position is an attractive one for two reasons. First, we might reasonably want to leave open the possibility that coercive action falls along a spectrum. This distinction can accommodate a spectrum approach without us needing to always demonstrate coercion-proper. Be that as it may, coercion and being coercive are related ideas. And we can only understand the latter by reference to the former. Even mere coerciveness demands that the coercive activity “provide a sufficient reason” for (not) acting in one way rather than another (Lammond 2000, 52). Thus, even if we allow that the pressure may not be overtly felt, two features are entailed. There needs to be a causative link to the intervention we are claiming is coercive and the effect of the intervention must not be easily resistible (even if it is not fully controlling in the manner of coercion-proper).

Secondly, the coercion-coercive distinction may accommodate and align more faithfully with one of the principal concerns of commentators regarding nudges, which is essentially about paternalistic action (Hausman & Welch 2010, Scoccia 2014, White 2013). White, for example, is concerned with the subversion (or substitution) of the true ends and interests of individuals by those of third parties, especially when they are purportedly underpinned by paternalistic justifications (White 2013, 64). In considering the cafeteria scenario, he asks how else the cafeteria choice architects could organise the food. He makes several suggestions including “common sense ordering . . . by weight or fragility . . . in an aesthetically pleasing manner”. He claims that the “important thing is that none of these plans takes into account any influence they may have on customer’s choices” (White 2013, 98). It would, of course, be very strange if cafeteria owners did not take account of the effect on customers of the layout of their premises. Presumably in their attempt to sell us food and beverages they are keenly interested in this. What White means by this comment is that these layouts do not purport to be in our own interests (i.e. to make us healthier). Nevertheless, while the motivations of the cafeteria choice architects may be germane to the question of coercion, it is not clear that this is anything to do with their beliefs (correct or not) about the interests of their patrons. As
we saw in section 3.2, intent can be highly relevant to our assessment of coercion, but it is intent to coerce in order to bring about a particular outcome which is key, rather than any intent regarding the coercee’s interests. If we want to connect coercion to a narrative about our interests (or will or autonomy), then it ought to be based on the effect of the intervention on those interests (notwithstanding the fact that we might exploit knowledge regarding a person’s interests in order to successfully coerce them). But nudges appear to be much more benign and uncertain in their effect than the infringement of interests which is suggested by paradigm cases of coercion. This is not to say that our own interests are not relevant in their own right to our assessment of the justifiability of nudges. Concerns of this ilk centre, in particular, on questions related to paternalism and whether paternalistic action or policy is defensible. It is entirely possible that paternalistic policies might be coercive, but coercion and paternalism are two separate issues. Coercion is not intrinsic to paternalism and vice versa (Thaler & Sunstein, 11-12, Conly 2013a, 242). If, therefore, we want to charge nudges with being coercive more is needed than simply appealing to their paternalistic nature (if indeed they are paternalistic).  

4. Concluding thoughts

We often use the word coercion loosely. As Anderson notes “one hears ‘coercion’ used to describe social pressures . . . ; or the constraining or manipulative effects of advertising, one’s upbringing, or the structuring of society more generally” (Anderson 2011, 2). This is all well and good so long as we do not mean to accompany such statements with any claims regarding the moral acceptability or not of such practices. Perhaps, therefore, claims of coercion and coercive action are not meant to be taken as literally or seriously as I have in this article. The purpose of such claims may instead be rhetorical, aimed at highlighting more important concerns about the use of nudges. We can see this in White’s comments set out in section two. A significant part of his unease regarding nudging centres on the substitution of the ends and interests of individuals with that of the nudger. Moreover, commentators often display a selective concern with attempts by government to employ the kinds of behavioural techniques under discussion. For example, although they concede that government attempts to influence citizens using these techniques is sometimes acceptable, Daniel Hausman and Bryn Welch say that “an organized effort to shape choices still appears to be a form of disrespectful social control” (Hausman & Welch 2010, 134). Nevertheless, coercion is a serious claim and is a high bar against which to measure the types of behaviourally-informed interventions at issue. As Lammond notes, it is “a particular means of altering human conduct whose evaluative significance depends on the nature of those means” (Lammond 2000, 46). Given this moralised overtone we need to be careful with our usage of the term.

21 Some might be, but, Thaler and Sunstein’s specific appeal to libertarian paternalism aside, this need not be a feature of all nudges.
Coercion, as we have seen, is not the same as merely steering people’s choices and decisions. It is patently qualitatively different. When we coerce someone we intentionally leverage some power or other that we have over them. We do this in order to compel them to act (or not) in a manner of our choosing rather than their own and the means by which we do this makes it exceedingly difficult for them to do otherwise. Whatever might be objectionable about attempts to harness or mitigate a person’s cognitive biases and effects to steer their behaviour or choices, such attempts are not generally or appropriately captured by charges of coercion. I do not discount the possibility that the choice environment or elements within it could amount to coercion, but I remain unpersuaded that the kinds of interventions under currently discussion in policy documents and the nudge literature do so. Crucially, they are not relevantly controlling, something which we saw in section 3.2 and 3.3 is implicated whether we theorise coercion on either the pressure or the enforcement approaches.

References


Caloia, B. Psychic costs and broad learning effects: Locating coercion and threats to autonomy in the use of ‘nudges’ in public policy. Pre-publication


Day, P. (1992) Coercive offers do not exist. Philosophical Notes, 24,


Hughes, R.C. (2013). Law and coercion. Philosophy Compass, 8, 231-240


Rozin, P., Scott, S., Dingley, M et al. (2001). Nudge to nobesity I: Minor changes in accessibility decrease food intake. Judgment & Decision Making, 6, 323-332


White, M.D. (2011). We’ve been nudged: The effects of the downturn on dignity and responsibility. (In M.A. Starr (Ed.), Consequences of economic downturn: Beyond the usual economics (pp. 103-118). New York: Palgrave MacMillan.)
