

Nudging and Manipulation

T. M. Wilkinson

University of Auckland

Behavioural economics and social psychology have shown that humans have all sorts of psychological quirks. Policy makers have become enthusiastic about taking advantage of these quirks through what Thaler and Sunstein call 'nudges'. This article asks: when would nudging be manipulative? The article has six parts: (1) publicity and transparency, which claims that Thaler and Sunstein's own attempt to deal with evil nudges is inadequate; (2) manipulation and autonomy, where the nature and wrongness of manipulation is connected to a conception of autonomy; (3) the perversion of the decision-making process – a piecemeal approach, which sorts nudges into easy and hard cases and assesses attempts to pick out certain methods, such as temptation, as manipulative; (4) the perversion of the decision-making process – general accounts, which shows why we do not have a clear, complete and correct account of what such perversion is; (5) intentions and nudging's escape clause, where it is shown that governments that nudge as Thaler and Sunstein would wish do not manipulate because they do not have the intention to manipulate; and (6) consensual manipulation, where it is claimed that manipulation can, with the right consent, be consistent with autonomy.

Keywords: nudge; manipulation; autonomy; social psychology; behavioural economics

Behavioural economics and social psychology have shown that psychological quirks routinely cause people to deviate in predictable ways from stereotypical economic rationality. Policy makers appear to be becoming enthusiastic about taking advantage of these quirks (e.g. Oullier *et al.*, 2009). Policies that go with the grain of people's psychology may be more effective than the stick of coercive regulation and cheaper than the carrot of subsidy. The best-known book that combines a synthesis of the academic literature with policy recommendations is *Nudge*, by Richard Thaler and Cass Sunstein (Thaler and Sunstein, 2008 – abbreviated to N in future references). Thaler is an economist, and now an adviser to the British government's Behavioural Insight team; Sunstein is an academic lawyer currently working for President Obama. Their term, 'nudge', is a memorable way of capturing the use of modern social science to steer people's behaviour.

The central idea of *Nudge* is that 'small and apparently insignificant details can have major impacts on people's behavior' (N, p. 3). Instead of making it illegal to sell junk food or taxing it to discourage consumption, people can be nudged into healthier behaviour, for instance by placing healthy food in a prominent position. Instead of the government's mandating the pensions or credit facilities it thinks would suit people best, it can improve their choice of existing options by making the best the default. Instead of (or perhaps as well as) large numbers of traffic police, the government can reshape the lines on the roads to slow drivers down (N, pp. 37–9). Instead of punishing litter droppers or turning off people's electricity if they use too much, governments or firms can encourage pro-social behaviour with sophisticated advertising (such as 'Don't mess with Texas', N, p. 61) or smiley faces on electricity bills (N, p. 69.) According to Thaler and Sunstein, these methods can be surprisingly successful in changing behaviour while being libertarian, in the sense that they leave people free to do what they want.



Several hard questions can be asked of nudging. Why think that nudges would be effective or make a substantial difference to social problems? Why think that nudging people would make them better off (Marteau *et al.*, 2011; Wilkinson, 2009)? I do not think these questions are unanswerable and I think 'nudging' is an imaginative approach to policy making which could turn out to be valuable. Nudging, however, raises a question that has not so far been much thought through. Is it not manipulation to take advantage of people's predictable deviations from economic rationality? And if it is manipulation, how can the nudging be libertarian?

Manipulation is a charged term. To be manipulated can be understood colloquially as being made someone's 'puppet on a string'. Such a simile seems overblown for some nudges, but not, perhaps, for all. For policy makers who want to change behaviour, nudging may seem benign. It is done for good motives, and it seems cheap and flexible and it leaves people free – what's not to like? By contrast, nudges seem to other people to be examples of the same sinister behind-the-back behaviour control that marketers and advertisers use in the private sector. Moreover, even if governments nudge out of good motives, the manipulative power used may be all the worse coming from a government that is supposed to represent its citizens and which, unlike in the private sector, is not counteracted by the efforts of competing firms. My own view is that neither of these views is correct; some nudging would be manipulative, and thus objectionable, and other nudging would not. This article tries to provide the basis for differentiating the manipulative from the non-manipulative.

The question of manipulation is important beyond Thaler and Sunstein's own work. Theirs is but one account of how to use social psychology and behavioural economics and how to understand those uses normatively. For instance, within a body of regulatory studies termed 'new governance', some writers suggest ways to use new insights to encourage firms' compliance with regulations. However defensible that may be, compliance with regulations differs from Thaler and Sunstein's nudges (Amir and Lobel, 2008, pp. 2132–7). Nor are the insights used only by governments or those who wish to promote social benefit. Many of the tricks of marketing count as nudges whether they are by firms trying to entice customers or governments trying to promote saving. However, although the discussion in this article is relevant beyond Thaler and Sunstein's work, theirs is, as I said, the best known and I take their book as a starting point in exploring the question of whether and when nudging is manipulative.

The article begins with Thaler and Sunstein's own account of nudging and their way of dealing with the problem of manipulation, which relies on publicity and transparency. In the usual essay marker's phrase, their way is fine as far as it goes, but they do not consider the conceptual and normative question of what it is to manipulate people and, without doing so, we cannot be confident that their solution is adequate. The article then puts Thaler and Sunstein aside so as to explore the difficult and under-considered concept of manipulation. This will be a long and, unfortunately, rather inconclusive discussion, which will show why one should not rush to judgement when trying to decide whether nudging manipulates. More conclusively, the remaining sections claim that nudges at least do not manipulate when they are based on the right intentions or when they are consented to. As we shall see, some nudges would manipulate, some would not, and some are hard to assess.

The article thus falls in between those who would detect no problem of manipulation with a nudge that leaves the option of opting out and those who see all nudging as manipulative.

Thaler and Sunstein's Response

Whether nudging manipulates depends in part on what nudging is. 'A nudge ... is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives' (N, p. 6). This definition is of limited help. How should one interpret 'choice architecture'? If a sign saying 'Danger! Thin Ice' causes people not to walk on ice, is the sign an aspect of 'choice architecture'? Is advertising, truthful or not, or the use of smell or colour, or any element of successful design? Some writers distinguish rational persuasion from nudging (Bovens, 2008; Hausman and Welch, 2010). For them, it is a defining feature of a nudge that it taps into the irrational or non-rational ways of thinking and acting that Thaler and Sunstein discuss. 'Danger! Thin Ice' would not then count as a nudge. However, I do not think the distinction between rational persuasion and using supposedly irrational or non-rational methods is at all clear even conceptually, let alone in practice, so I doubt the distinction helps in defining a nudge. In fact, *Nudge* says little at the conceptual level about the concept of a nudge and readers have to draw their inferences from the numerous examples Thaler and Sunstein provide and from how they use behavioural economics and social psychology.¹ The inference I draw is that Thaler and Sunstein's nudging differs from rational persuasion in relying on modern social science's insights into human thinking and behaviour, and nudges include any methods, apart from carrots or sticks, that rely on those insights.

Thaler and Sunstein do not consider in any serious way the conceptual question of whether nudging is manipulative. They come closest when they briefly consider what they call 'evil nudges', such as subliminal advertising, and the problems of drawing the line between nudges that should and should not be permitted. Thaler and Sunstein hope to avoid evil nudges through publicity and transparency, which I shall discuss after a brief comment on the inevitability of nudging.

Thaler and Sunstein stress throughout their book that options often have to be presented in some way that inevitably shapes choices. Some food item has to come first at the cafeteria. Some pension plan or none has to be the default. When shaping choices is inevitable, they say, it is often best to present them so that people are more likely to end up with the option that is better for them (e.g. N, pp. 2–4, p. 243, p. 247). However, the inevitability argument is not as forceful as Thaler and Sunstein think. If, on the one hand, the methods are otherwise objectionable, an opponent of manipulation could well say that, contrary to Thaler and Sunstein, options would better be presented at random than with the intention of steering behaviour. Then no-one would be under anyone else's control. If, on the other hand, the methods are not objectionable, then the inevitability argument is redundant. Thaler and Sunstein need not cite inevitability in defence of methods to which there is no independent objection. (The argument for getting people to choose what is best has some force when objectionable methods are unavoidable – but why should they be?)

I come now to Thaler and Sunstein's publicity principle and their aim for transparency. They write: 'In its simplest form, the publicity principle bans governments from selecting a policy that it would not be able or willing to defend publicly to its own citizens' (N, p.

244). A requirement of publicity would constrain government actions in practice because objectionable nudges would cause them embarrassment. And a publicity principle is ethically required by a duty to show respect to citizens.

The publicity principle clearly needs elaborating. Thaler and Sunstein explain that, when they consider the defence of a policy, they mean both the end that the policy aims at and the methods it uses; but they do not explain what they mean by being 'able' to defend the policy publicly. It cannot mean, in ordinary language, that governments can *successfully* defend the policy, and it cannot mean that the policy is right.²

The publicity principle is fine as far as it goes. What if governments were so confident in their wisdom that they would be able and willing to defend a method we would see as clearly objectionable (and manipulative)? To take Thaler and Sunstein's example, what if a government were to use subliminal advertising to reduce excessive drinking and then give a stout pro-public health argument for doing so (N, p. 246)? Thaler and Sunstein produce what appears to be another reason to oppose such a method: manipulation of this kind is objectionable precisely because it is invisible and thus impossible to monitor (N, p. 247).

Their reason is set out in only a sentence, so we are not told who it is who is supposed to be doing the monitoring and how the impossibility of monitoring makes a nudge objectionable. It could be that nudges are objectionable when the nudged themselves cannot monitor the process. Or perhaps it would be enough that informed critics could monitor a process and make their results known to the public, thus forcing governments to live up to the publicity principle.

One way to interpret Thaler and Sunstein's few comments on evil nudges is that they are doing political theory. They are confronting worries about giving newfangled power to malicious or overbearing governments by recommending (sketchily described) publicity and transparency constraints on the exercise of that power.³ If this interpretation is right, political theory is fine, but not enough. Maybe a government committed to publicity and transparency could not use nudges that are objectionably manipulative. But we cannot be confident that publicity and transparency in combination would take away a government's motive and opportunity to manipulate unless we have a reasonably clear understanding of what it would be to manipulate.

Manipulation and Autonomy

As much of the article will try to show, it is a surprisingly difficult question whether and when nudging is manipulative, because the concept of manipulation is itself surprisingly difficult to formulate and apply. People can be manipulated when they go shopping, strike contracts, vote, study at school, visit their doctors, decide whether to have sex or take turns to do the housework. A full account would have to cope with the enormous variety of sites and methods of manipulation. Indeed, we do not have such an account. A diverse literature in psychology, communication studies, marketing and advertising contains a wealth of insights into methods of manipulation.⁴ But they are yet to be integrated into a satisfying normative theory. There is no equivalent for manipulation of the monographs written on exploitation, coercion, liberty or autonomy.

However, to get a sense of what is involved in the concept, consider Joseph Raz's view: 'Manipulation, unlike coercion, does not interfere with a person's options. Instead it

perverts the way that person reaches decisions, forms preferences or adopts goals' (Raz, 1986, pp. 377–8). For Raz, manipulation is wrong, except to prevent harm to others, because it violates people's autonomy (Raz, 1986, p. 378, p. 420). Thomas Hill writes: 'Manipulation, broadly conceived, can perhaps be understood as intentionally causing or encouraging people to make the decisions one wants them to make by actively promoting their making the decisions in ways that rational persons would not want to make their decisions' (Hill, 1991, p. 33). For Hill, too, manipulation is objectionable because it violates autonomy. Putting these views together, we can say the following as a first approximation to saying what manipulation is, why it tends to be wrong and why one might think that nudging is manipulative.

First, manipulation involves the perversion of a decision-making process. Whereas coercion uses threats, which involve changing the costs of selecting certain options, manipulation involves some underhand interference with the ways in which people see their options. What counts as a perverted decision-making process is something we shall come to, but we can immediately dismiss the idea that an autonomous choice must be free from the influence of other people. That idea would imply, among countless bizarre results, that you would manipulate me if you shout a truthful warning not to cross an unsafe bridge and I, hearing your warning, decide not to cross it. Some influences are consistent with autonomy; others are not. The difficulty, which is also a large part of the difficulty in understanding manipulation, is to explain and justify the line between them.

Second, note that manipulation involves an intentional actor. This intention condition is of great importance in explaining why manipulation is wrong. Take lying as a canonical method of manipulation. A person may have false beliefs about his or her options without his or her autonomy being affected; who has true beliefs about all their options? But if those beliefs came about through deceit, his or her autonomy has been harmed. From the point of view of autonomy, it matters how we come to our beliefs and desires. In this respect, manipulation is similar to coercion. People can be autonomous while choosing from a limited menu of options; autonomy cannot require choice from an unlimited menu. But if an option is made ineligible because of coercive threats, our autonomy is infringed upon (Raz, 1986, pp. 377–8). The extent of manipulation is a matter of degree but, even if it falls short of full control, manipulation is a form of influence that subverts and insults a person's autonomous decision making (Hausman and Welch, 2010, p. 128, p. 133; Mills, 1995, pp. 106–7, has a contrary view).

Third, consider some points about the wrongness of manipulation. What is primarily wrong about manipulation is that it violates autonomy. Manipulation could be wrong for other reasons, for instance because it causes us to be exploited or to act against our interests, but manipulation would be *prima facie* objectionable even if it made people better off. This view is widely shared. Although Thaler and Sunstein want their nudging to be 'libertarian', a word that usually denotes a narrow and extreme anti-state political doctrine, the sense that people should not be the puppets of others is to be found in liberal thought and Marxian and feminist critiques of false consciousness. To manipulate people is to treat them as both tools and fools. So I shall take it that it would be a major drawback to a nudge if it was manipulative, a view that Thaler and Sunstein would no doubt share, given their aim of showing nudging to be libertarian.

That said, manipulation could be justified even if it does violate autonomy. Even if redrawing lines on roads did constitute manipulation, it would surely amount to a very small loss of autonomy and, if it saved lives, it would be worth that loss. Manipulation could perhaps be justified if it were designed to counter other forms of manipulation. Exaggerating the harms of alcohol is sometimes defended as counteracting the better-funded advertisements of alcohol producers. So to show that nudging was manipulative would not be to rule it out. But it would make nudging hard to justify and indeed unjustified in many cases.⁵

Although I have said that manipulation infringes upon autonomy, this claim is too crude. As a later section points out, manipulation could be consented to. If it were consented to, in the right kind of way, then the manipulation would at least be consistent with autonomy and might count as enhancing it.

Fourth, Thaler and Sunstein are interested in what they call 'libertarian paternalism': nudging as a way of steering people's choices in directions that will improve their lives without restricting their freedom. However, as is shown by the examples of discouraging speeding, littering and using too much electricity, they are not considering paternalism alone. They also want to consider methods that will steer people into acting so as to benefit or not harm other people. It may be thought that manipulation that is not paternalistic could not violate autonomy. Autonomy is about self-rule, and the boundaries of self-rule do not include dropping litter or speeding. However, we surely believe that duplicitous political propaganda can manipulate voters by stopping them making up their own minds and thereby infringing on their autonomy; and we hold this belief even though the political matters need not be matters of self-rule (Scanlon, 2003). So concerns about manipulation apply to non-paternalistic nudges as well as paternalistic ones.

Finally, I mention a difficulty with telling whether someone is manipulated. Manipulation requires that the target decide *because* his or her decision is *perverted* by the manipulative act, so telling whether someone has been manipulated requires telling whether that person's decision was both caused by the manipulative act and in some perverted way. As we all know (although often forget in practice), correlation is not causation, so showing causation satisfactorily requires some effort. Marketers do make such efforts, when assessing the influence of their advertisements, and I do not want to argue that it is impossible or even impractical to demonstrate causation. Doing so does, however, require more than pointing out that, following a manipulative act, someone acted in the way the manipulator desired. As an added difficulty, one would have to show that the causal route was a perverted one, however it has been decided what constitutes a perverted route. Whether a decision was perverted may well not be obvious. For example, suppose people are exposed to direct advertising for pharmaceuticals and subsequently ask their doctors about them. On one view, they have been manipulated by misleading advertisements that imply near-miracle results. But on another view, they have discounted the puff, in the way many people do with advertisements (Offer, 2006, pp. 121–2; Vaithianathan, 2006), but noted that the pharmaceutical exists and can be used for their condition, which will be true. On this rosier interpretation of advertising, the advertisement communicates facts, whatever else it does, and if people act on those facts, they may not be manipulated. As with causation, it may be possible to test people's reasons for acting to decide whether their decision was perverted.

The point here is that people may not be manipulated even if they do what the ‘manipulator’ wanted as a result of the ‘manipulator’s’ action.⁶

The discussion in this section boils down to this: manipulation is intentionally and successfully influencing someone using methods that pervert choice. The remaining sections discuss the implications for nudging which are, in brief, as follows. Nudging clearly could satisfy the condition of having an intentional actor who successfully influences someone’s decisions. In principle, nudging could also pervert someone’s decision-making process and thereby infringe upon his or her autonomy. But if and when nudging does not use a perverting method, it is not manipulation. If and when nudging is not a form of intentional influence, it is not manipulation. If and when nudging does not succeed in altering behaviour, it is not manipulation. Finally, if the target consents in the appropriate way to nudging that would otherwise be manipulation, the nudging would not infringe on the target’s autonomy.

The Perversion of the Decision-Making Process: A Piecemeal Approach

The next two sections are about the perversion of the decision-making process, an idea at the heart of the concept of manipulation. ‘Perversion of the decision-making process’ is an abstract concept. One can take a piecemeal approach to deciding what nudges fall under it by considering cases or trying to pick out the objectionable methods without trying to say what they have in common to make them such a perversion. This section describes a piecemeal approach and its limits. The next considers attempts at a general account of manipulation. As I shall explain, we have no clear, complete and correct view of what constitutes a perversion. The upshot is that it will be hard to decide whether many nudges are manipulative. This conclusion is an important one because, at least in discussion, nudges are often thought – too quickly – to be manipulative.

Let us begin with some easy cases. Some cases are clearly manipulative. Successfully deceiving someone by lying is one canonical example of manipulation. So too, if it worked, would be subliminal advertising. Since lying and subliminal advertising are not excluded by the definition of a nudge, we can say that some nudges could be manipulative. Thaler and Sunstein must agree; if a nudge by definition could not manipulate, why set out publicity and transparency conditions to prevent manipulative nudges?

Some cases are obviously not manipulative, for instance warning of an unsafe bridge. Can we find nudges that do not manipulate? Thaler and Sunstein give an example with the acronym RECAP (Record, Evaluate and Compare Alternative Prices). RECAP uses the knowledge gleaned from social psychology and behavioural economics to present information in a way that stands out. Among other applications, Thaler and Sunstein suggest using RECAP for credit card expenses (N, pp. 143–4). Suppose many people, once acquainted with the facts, would spend less on their credit cards and suppose the nudger who uses RECAP intends that outcome. The nudger’s aim is to alter behaviour in a predictable (and desired) way using fancy nudge techniques. And yet the person who then spends less does so because he or she better understands what is at stake. So RECAP seems non-manipulative. Or consider scaring people with relative risk (‘your chance of a stroke is doubled if you don’t take these pills’) (Moynihan and Cassels, 2005, pp. 84–6). Suppose psychologists and doctors want to avoid needlessly scaring patients, and take considerable

pains to present probabilistic information so that patients can understand it and appreciate its significance. Suppose their aim is to prevent patients demanding unnecessary medication for high blood pressure by showing them that their marginal risk without it is in fact low. The method is not simply straightforward presentation of facts, and it uses knowledge of psychology, so it should be classed as a nudge, but intuitively it seems non-manipulative. (The next section gives some theoretical reasons to think that RECAP and the blood pressure example of risk framing are non-manipulative.)

Other cases are harder. Lying may be manipulative, but is failing to disclose all the information? What if the non-disclosure were to avoid giving away intellectual property? Subliminal advertising may be manipulative, but is playing music in restaurants to encourage people to eat more? What if the music worked because people feel uncomfortable eating in a silent restaurant (even if they do not realise that they do)? If advertisers use colour to make their products stand out and more people buy them as a result, is that manipulative? If they use a comedian to tell jokes in their beer advertisements, is that manipulative, and any more so than using humour in teaching? Some people resent marketing tricks, when they are pointed out, such as placing a particular brand of beer in a superficially random but highly visible pile in the supermarket. Other people, who want to economise on their shopping time and do not much care what beer they buy, may not mind; making that beer visible means they do not have to spend time looking. People may resent sweets at the checkout, designed to stimulate impulse buying, but that is probably more to do with the temptation of sweets than the visible placing; I doubt that people feel comparably irritated by a visible display of batteries.

Perhaps we could identify certain methods as manipulative, for instance those that play on emotions, or tempt us to act against our better judgements. Emoticons on our electricity bills, aimed at making us feel morally superior or inferior to others in our environmental habits, could thus count as manipulative. But it cannot be manipulative in all cases to use a method that relies on our emotions. Suppose I believe I should install smoke alarms in my house, but do nothing about it until I see a vivid advertisement about the dangers of fire. The fear the advertisement arouses gets me to do what I already thought I ought to do. We can welcome the use of our emotions to turn our pale beliefs into vivid ones and act as we would, on reflection, want.

Temptation is different. It is plausibly part of the meaning of temptation that, when we act because we are tempted, we act contrary to our better judgement. If the temptation has a tempter, it is plausibly part of the meaning of tempting someone that one does so intentionally. A person tempted by a tempter sounds manipulated. Tempting implies playing on a person's weakness, which some have identified as a key element in manipulation (Rudinow, 1978). However, determining whether a nudge is temptation may not be easy. Recall the example of the beer display, which shows that some acts may be intended to tempt and may cause the desired action and yet not be manipulation if, for instance, the person was pleased to have the reminder to buy more beer. Other acts may tempt someone but not be intended to tempt and would not count as manipulative temptation.

This section shows that a piecemeal approach to thinking about the perversion decision making will get us some way when deciding whether a nudge is manipulative. However, some nudges are already hard to assess and since nudging is, or at least is hoped to be, a

continuing programme, many more such nudges are likely to be invented. Trying to think in a nuanced and contextualised way whether a nudge is manipulative may leave us at a loss. A general account of manipulation would then be desirable. The next section considers the prospects for a general account. Unfortunately, they are not bright.

The Perversion of the Decision-Making Process: General Accounts

When trying to decide what perverts decision making, one suggestion is to consider people's feelings subsequent to acting on some agent's influence. If they feel they have been manipulated, that is a reason to think they have been manipulated; if not, it is a reason to think they have not been. This suggestion may well have an advantage in suggesting a practical test for manipulation: ask people how they feel, and perhaps generalise from them to others in their position. The suggestion does not, however, help us with analysing the concept of manipulation. First, people may not notice that they have been manipulated. Indeed, manipulation is often thought to work best when the victims are not aware of it. So a feeling of manipulation is not necessary for manipulation. Second, people may incorrectly feel manipulated. This claim is clearly true of lying as a form of manipulation; one can *feel* lied to without having been lied to. So a feeling of manipulation is not sufficient for manipulation. Indeed, to use people's feelings as a test for whether manipulation has occurred could only be validated once we were clearer about the concept of manipulation.

Some attempts at explaining manipulation ask about the quality of the decision. Reconsider the cases of RECAP and risk framing, which I claimed were non-manipulative nudges. We could judge these nudges not to involve manipulation for at least three reasons. It could be said (1) that RECAP and the risk framing improve the recipients' appreciation of the relevant reasons that apply to them and the decisions they make are better according to some theory-derived standard. It could be said (2) that the decisions are not manipulated because the recipients themselves agree that they have a greater appreciation of the relevant reasons. The decisions they make are better according to their own standard. And it could be said (3) that the decisions are better according to a standard endorsed by those doing the framing, which is what would matter on a view that takes manipulation to be defined entirely in terms of the manipulators' beliefs and intentions (see, e.g., Noggle, 1996). All three explanations agree that RECAP and the risk framing are not cases of manipulation.

In the rest of this section I consider the difficulties that face the first two explanations, which conceptualise the perversion of a decision-making process in terms of a theory-derived standard and the standards of those making the decision. I do not consider the third explanation. Although, as the next section shows, we can learn much about manipulation from the intentions of the manipulators, it is too implausible to think they are the whole story.

Consider the earlier quotation from Thomas Hill, where he defines manipulation as 'intentionally causing or encouraging people to make the decisions one wants them to make by actively promoting their making the decisions in ways that rational persons would not want to make their decisions'. Hill's suggestion for identifying manipulation asks us to think about rational people making a meta-decision: a decision about their decision-making process. Specifically, he asks us to think about what rational people would identify as flaws in such a process, such that being intentionally caused to decide through those flaws

would constitute manipulation. On one interpretation, rational people would want to decide in an ideally rational way, for example, deciding in the light of full information, choosing the most cost-effective means to pursue given ends, with options ranked in some consistent preference ordering, and so on. A manipulated decision would be one that has been caused to fall short in some dimension of a fully rational choice.

Selective disclosure, burying the information in small print, framing risk information to scare patients, getting people drunk before they decide; these could all be picked out as manipulative because they involve some people causing or encouraging other people to decide in a way that falls short of an ideal decision-making process. This suggestion may appear obviously wrong in requiring decision making to be hyper-rationalistic if it is to be autonomous. People's choices may fall short of some idealised model of rational decision making and yet be voluntary enough (Feinberg, 1986, pp. 104–5, pp. 113–7). However, the suggestion is not wrong for that reason. Even if less than fully rational decisions may be voluntary, deliberately causing someone to fall short of full rationality could be manipulation. For instance, while proceeding with some transaction in the absence of full information may be voluntary enough, the deliberate non-disclosure of some fact by the counterparty could be manipulation.

Even so, the suggestion that manipulation is intentionally causing or encouraging people to decide in ways contrary to an ideally rational process cannot be correct. The root of the problem is that the ideally rational process is not in fact rational for us. Many writers have stressed the ways in which rational creatures, under pressure of time and facing the costs of gathering information, will find short cuts when making decisions. Thaler and Sunstein point out some of the ways in which these short cuts can go wrong and other writers point out the ways in which past short cuts are not well suited to an affluent world (for instance, having a primitive fear of the dark but not of smoking tobacco). Nonetheless, rational people will often not want to decide in well-informed ways following careful consideration, and so on. These methods can themselves yield the wrong answers, or the right answers, but too late (Dennett, 1984; Gladwell, 2005).

Hill did not explicitly say that rational people would want to make their decisions in an ideally rational way. But how, then, would rational people want to make their decisions? Rational people would care both about the process itself and the outcome but they might disagree – and still be rational – about the relative weights to give process and outcome. Suppose a nudge causes a time-pressured person with limited attention to make a decision a certain way, and it is in that person's interests: for instance, enrolment in a pension plan. One rational person may oppose making a decision in such a way, perhaps preferring in principle to decide free of that kind of intentional influence or perhaps believing, as a matter of policy, that nudgers cannot be trusted. But another rational person may reason that it is better to get the right result, in this case enrolment in the pension plan, and welcome the nudge. When it comes to deciding which difficult cases are ones of manipulation, the answer will not be given by asking what rational people would oppose because rational people would not agree with each other.

A theory-derived standard has the drawback of being either question begging – why that standard when many would not accept it? – or inconclusive. What if we were to use the recipients' own standard for decision making? As before, the idea is that we seek a

meta-decision, a decision about their decision, but this time we use people's own standards, not a theory-derived standard such as ideal rationality. This idea may seem to have the drawback of being inconclusive again. A given nudge may come out as manipulative on some people's standards but not manipulative on others. However, it would be more accurate to say of such a nudge, if we are using the recipients' own standards, that it would manipulate some people but not others. That is more conclusive than a rationality standard that is unable to judge whether anyone was manipulated.

While the meta-decision that uses people's own standards may prove conclusive, it is subject to an apparently insuperable problem. If we are trying to decide whether a decision has been manipulated, asking people whether they endorse the decision they have made presupposes that their endorsement has not itself been manipulated. The problem of deciding on manipulation thus recurs at the meta-level (Sumner, 1996, pp. 168–9). Some writers have attempted to find a way around this problem. In effect, they say that not any endorsement counts, but only endorsement that meets certain conditions, for instance one made in the light of certain facts. The problem then is this: if we know what conditions would make the meta-decision non-manipulative, then we already know the conditions that would make the original decision non-manipulative, in which case we do not need the idea of a meta-decision at all. But we do not have a complete account of these conditions and we cannot rely on a meta-decision that presupposes such an account.

Let me summarise this and the preceding sections. The question is whether and when nudges would be manipulative. In order to be manipulative, nudges would have to pervert the decision-making process. We do not have a good general account of that perversion. What we have are some intuitions about certain methods, such as lying and temptation, which could be applied to some nudges. We also have some cases where our intuitions give the same answer as several of the extant (and inadequate) general accounts of manipulation, as in the cases of RECAP and risk framing, which may give us more confidence in our judgement of them as non-manipulative. At this stage, there seems no alternative to a piecemeal contextual approach in trying to judge whether a nudge perverts decision making. However, such perversion is only one element in manipulation. When we consider intentions and consent, we will be able to draw more general conclusions about nudges.

The Intention to Manipulate and Nudge's Escape Clause

I turn now from the methods of manipulators to their intentions. Manipulation involves intentionally getting someone to do something. Manipulation infringes upon the autonomy of the victim by subverting and insulting their decision-making powers. While nudging is intentional, the question arises whether the intentions behind nudging amount to subversion or insult. Nudges by definition leave people free to opt out and Thaler and Sunstein hope people will opt out of nudges that do not suit them personally or are poorly designed. Their nudges have what I shall call an 'escape clause'. When nudges have an escape clause, I shall argue, they are not manipulative because they do not satisfy the intention condition.

Thaler and Sunstein claim that theirs is a libertarian form of paternalism because people are free to opt out. They want the opt-outs to be genuinely available because they want people to opt out of nudges that do not work for them. For them, part of the attraction of

nudging, as a method of paternalism, is that it better meets the objection ‘how do you know what is good for people?’ than traditional command and control paternalism. Traditional regulation is less sensitive than nudging to the differences in people’s interests (e.g. N, pp. 186–7). Furthermore, governments, like anyone, can make mistakes. Opt-outs play an important role in correcting for government error. If a government’s nudges are misguided, whether paternalistic or aimed at the social good, people can opt out. The freedom to opt out is not one that Thaler and Sunstein regard as merely formal, something to be tossed in the direction of libertarian critics to assure them that the nudged still have some freedom.

Suppose a government really wants people to opt out of nudges that the nudged themselves find misguided or unsuitable. In other words, suppose the government wants people to be genuinely free to opt out whether for good or bad reasons. I believe this goes a long way towards showing that their nudges would not be manipulative because the government would not have the kind of intention that manipulation requires. These nudges would not take the form ‘A intends B to do X’ but rather ‘A intends B to do X unless B wants not to’. There is certainly an intention there, but it is not the imposition of the manipulator’s will on someone else. Consequently, I think nudgers who sincerely want people to opt out of unsuitable nudges are not imposing their will, so not manipulating. I do not know whether this claim rests on a controversial view of manipulation, but the sources cited earlier connect intention to manipulation in the same way that I have done.⁷

Suppose the escape clause does not work. Suppose that, while people have the formal freedom to opt out, in fact the nudging methods are so powerful that people go in the direction they are nudged and do not opt out when the nudges are poorly designed or do not suit them. The argument could then be made, against the nudger, that while the nudger may not *intend* the result, and so may not count as a manipulator for the reasons just given, the nudger is being reckless or negligent. If we go back to autonomy, which underlies our concern about manipulation, the anti-nudge argument could go like this: people are being steered a certain way as the result of the intentional actions of someone else. They do not see the options they have. True, the nudger did not intend them to fail to see their options, but if a nudger does not care (recklessness) or ought to have known better (negligence), the nudger has come close enough to manipulation to count as infringing on autonomy.

Nudging could cause people to fail to opt out in cases where they should have done. If nudgers do not care or should have taken steps to prevent it, they may well have done something wrong. So much can be conceded to the argument in the previous paragraph. However, we are considering a nudging government that does sincerely want people to opt out, so the government cannot be reckless, by definition. Sincerely wanting an opt-out implies wanting people to have a genuine choice, which in turn implies avoiding methods that bypass choice. Nor need a government be negligent, if it makes genuine efforts to find out what its nudges do to and for people. So I believe we can conclude the following: nudging is not manipulative when a nudger sincerely intends people to act on the escape clause and takes steps to find out whether people could.

However, we should also bear in mind the following conclusion. A government would satisfy the intention condition of manipulation if, when it nudges, it sees the escape clause as purely formal and hopes people will not think hard enough to opt out. If the methods were also manipulative and if they worked, then such a government’s nudges would be

manipulation. The point of this section is that nudges without manipulative intentions would not be manipulative.

Consensual Manipulation and Nudges

People sometimes play tricks on themselves to overcome the likely defects in their decision making. They may put their cigarettes where it is hard to get them, move their food away from their televisions or put reminder notes around the house. There are limits to what people can do to themselves, however. If they set the alarm clock for their early morning run, they can always turn it off when it rings (Schelling, 1984). So people may welcome third-party assistance. While this assistance could take many forms, one may be manipulation. If the manipulation is consensual, as this section argues some could be, then it is consistent with a person's autonomy. Nudges that are manipulative but consented to would then be consistent with autonomy.

One puzzle is whether we should think of consenting as making a manipulative act not wrongful or making the act not manipulative at all. If I say to you 'Don't bother me with the details' and you do not, the natural interpretation is that I have delegated to you and your not mentioning some information is not manipulative at all. The reason is that you have no manipulative intention. In other cases, such as consenting to be lied to, consent prevents manipulation. Once the scales have fallen from one's eyes, they cannot simply be put back again.

However, perhaps some acts could be manipulation despite being consented to. Suppose I agreed to be a subject for a stage hypnotist, and suppose, as popularly believed, hypnosis makes the hypnotised person act non-voluntarily. If I cluck like a chicken on the hypnotist's orders, this does seem to me a case of manipulation, although not a wrongful type because of the consent. If anything is like being a human puppet, it is hypnosis (in the folk version) and puppetry is the classic manipulation. The stage hypnotist example, if it is accurate, would be one example of consented-to manipulation. Other examples could occur when people agreed in advance to be shielded from certain facts or certain environments, as for instance with some addicts who wish to avoid cues that would trigger relapses. These are only possibilities and consent actually may make attempts to manipulate fail. It is an empirical question to what extent consent would make attempted manipulation fail and I do not know the answer. I can say that that extent will depend on what we take consent to require. Suppose, for instance, that retrospective consent could legitimate an act. In that case, since the information needed for consent would be given after the manipulation, it would not prevent the manipulation succeeding. Suppose we take a public vote as good enough.⁸ Some people may not have realised there was a vote, so they would not have the knowledge to make the manipulation unsuccessful.

As for what sort of consent should be required, the answer depends on the importance of individual autonomy and the extent to which people may autonomously limit their own autonomy. I have the space only to assert dogmatically that I believe people's autonomy does entitle them to limit their own autonomy to some extent, a view that is not especially controversial.⁹ But I also believe that individuals have to waive their own autonomy, so that a democratic choice does not count as waiving autonomy. To make this point clearer, I am considering whether an act of manipulation can be made consistent with autonomy by

consent and asserting that the answer is yes, but only when the individual consents and not when other people do, in a democratic majority or not. Suppose, for example, a manipulative nudge to discourage smoking were argued for on the grounds that most smokers want to give up. I am saying that the minority of willing smokers would have their autonomy infringed upon by the manipulative nudge (and so would those of the majority who did not consent to it). While the nudge could still be justified, by democratic choice or for some other reason, the nudges would come at the expense of the autonomy of those who did not consent.

Conclusion

People seem not to make their decisions in accordance with stereotypical economic rationality, and that must make some difference to the decision situation that governments or others ought to put them in. We should not oppose all attempts to take account or even advantage of people's psychology. But nudges can manipulate even if the aim is to benefit the victim or society and even though, by definition, they come with the formal freedom to opt out. The question of this article was how to decide which nudges are manipulative infringements on autonomy. I set out various conditions that can be used to determine whether a nudge is manipulative. A nudge would be manipulative only if the method perverted the target's decision. We do not have a good general account of such perversion but a piecemeal approach allows us to draw conclusions about specific nudges. A nudge would be manipulative only if the nudger had manipulative intentions. A nudge does not meet this necessary condition if it has a genuine escape clause, that is, if the nudger sincerely wants the targets not to act in the nudged way if the nudging is unsuitable for them. And a nudge may be manipulative but consistent with its target's autonomy if, but only if, the target consented to it.

Possibly none of Thaler and Sunstein's favoured nudges would be manipulative, assuming they consistently follow their own recommendation for an escape clause. However, they are not the only ones who are enthusiastic about taking advantage of psychological quirks. Among others, governments that want cheap behaviour change, health professionals focused on improving health, and advertisers and marketers who want to sell products, may well have manipulative intentions and use nudges that pervert the decision process. If those nudges are not consented to by their targets, they would manipulate and infringe upon the target's autonomy. Such nudges will be hard to justify.

(Accepted: 3 November 2011)

(Published online: 7 September 2012)

About the Author

T. M. Wilkinson is an Associate Professor in Political Studies at the University of Auckland. He has published widely in political theory and applied ethics, most recently a book called *Ethics and the Acquisition of Organs* (Oxford University Press, 2011). He was Chair of New Zealand's Bioethics Council 2006–9. T. M. Wilkinson, Department of Political Science, University of Auckland, Private Bag 92019, Auckland, New Zealand; email: m.wilkinson@auckland.ac.nz

Notes

My thanks for helpful discussion to Theresa Marteau, Stephen Winter, three anonymous referees and audiences at the University of Amsterdam, Keele University, University College London and the University of York.

- 1 Unfortunately for the task of understanding nudges, Thaler and Sunstein occasionally imply that policies they like are nudges even though they should not count as nudges on their definition because they do change incentives or ban activities. Cap and trade to control carbon emissions is one example (N, p. 187) and gambling self-bans another (N, p. 233). While we are on the subject of their conceptual shortcomings, Thaler and Sunstein peculiarly count any policy as paternalistic 'if it tries to influence choices in a way that will make choosers better off' (N, p. 5), which would make even rational persuasion paternalistic. And they have coined the term 'libertarian paternalism' to describe the broad thrust of their ideas but do not explain how libertarian paternalism relates to nudging. The answer seems to be that some nudging is libertarian paternalism, but some nudging is not paternalistic, some is not libertarian, and some libertarian paternalism is not nudging.
- 2 Hausman and Welch (2010, p. 135) point out that being able or willing to defend a policy does not explicitly require actually publicising it. However, the surrounding text in *Nudge* makes it clear that Thaler and Sunstein envisage actual publicity.
- 3 The objections they are responding to can be found in Glaeser (2006).
- 4 Cialdini (2007) is particularly well known and valuable.
- 5 For further discussion of the ethics of nudges, see Bovens, 2008; Hausman and Welch, 2010.
- 6 A question does arise about how to judge failed attempts at manipulation. In general, failed attempts to do wrong are nonetheless blameworthy but, in this case, some allowance should perhaps be made for the benefit of having some manipulative attempts counteract others (Mills, 1995, p. 211). Note, though, that one does a nudge no favours by arguing that it would not manipulate because it would not work.
- 7 I set aside difficulties with attributing intentions to complex organisations such as governments. Perhaps, if governments cannot have intentions, they cannot have the intentions necessary for manipulation.
- 8 Feinberg, 1986, pp. 66–7; Offer, 2006, pp. 72–3. Feinberg clearly envisages consent as individual informed consent; Offer seems to think democratic choice would be good enough.
- 9 See Feinberg, 1986. See also Scanlon (2003), who endorses a voluntary agreement not to be told unwanted facts.

References

- Amir, O. and Lobel, O. (2008) 'Stumble, Predict, Nudge: How Behavioral Economics Informs Law and Policy', *Columbia Law Review*, 108 (8), 2098–138.
- Bovens, L. (2008) 'The Ethics of Nudge', in T. Grüne-Yanoff and S. O. Hansson (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, pp. 207–20.
- Cialdini, R. (2007) *Influence: The Psychology of Persuasion*, revised edition. New York: Collins Business.
- Dennett, D. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge MA: MIT Press.
- Feinberg, J. (1986) *Harm to Self*. New York: Oxford University Press.
- Gladwell, M. (2005) *Blink: The Power of Thinking without Thinking*. London: Penguin.
- Glaeser, E. (2006) 'Paternalism and Psychology', *University of Chicago Law Review*, 73 (1), 133–56.
- Hausman, D. and Welch, B. (2010) 'To Nudge or Not to Nudge?', *Journal of Political Philosophy*, 18 (1), 123–36.
- Hill, T. E. (1991) *Autonomy and Self-Respect*. Cambridge: Cambridge University Press.
- Marteau, T., Ogilvie, D., Roland, M., Suhrcke, M. and Kelly, M. (2011) 'Judging Nudging: Can Nudging Improve Population Health?', *British Medical Journal*, 342, 263–5.
- Mills, C. (1995) 'Politics and Manipulation', *Social Theory and Practice*, 21 (1), 97–112.
- Moynihan, R. and Cassels, A. (2005) *Selling Sickness: How Drug Companies are Turning Us All into Patients*. Crows Nest NSW: Allen & Unwin.
- Noggle, R. (1996) 'Manipulative Actions: A Conceptual and Moral Analysis', *American Philosophical Quarterly*, 33 (1), 43–55.
- Offer, A. (2006) *The Challenge of Affluence*. Oxford: Clarendon Press.
- Oullier, O., Cialdini, R., Thaler, R. and Mullainathan, S. (2009) 'Improving Public Health Prevention with a Nudge', in O. Oullier and S. Sauneron (eds), *Improving Public Health Prevention with Behavioural, Cognitive and Neuroscience*. Paris: Centre for Strategic Analysis, pp. 38–46.
- Raz, J. (1986) *The Morality of Freedom*. Oxford: Clarendon Press.
- Rudinow, J. (1978) 'Manipulation', *Ethics*, 88 (4), 338–47.
- Scanlon, T. (2003) 'A Theory of Freedom of Expression', in T. Scanlon (ed.), *The Difficulty of Tolerance*. Cambridge: Cambridge University Press, pp. 6–25.
- Schelling, T. (1984) *Choice and Consequence*. Cambridge MA: Harvard University Press.
- Sumner, L. W. (1996) *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- Thaler, R. and Sunstein, C. (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven CT: Yale University Press.
- Vaithianathan, R. (2006) 'Better the Devil You Know than the Doctor You Don't: Is Advertising Drugs to Doctors more Harmful than Advertising to Patients?', *Journal of Health Services Research and Policy*, 11 (4), 235–9.
- Wilkinson, T. M. (2009) 'Reason, Paternalism, and Disaster', *Res Publica*, 15 (2), 203–11.