THE JUSTIFICATION FOR NUDGING

For Your Own Good: Informing, Nudging, Coercing

JUDITH LICHTENBERG*

INTRODUCTION

It is not news that people do not always act in their own best interests. They fail to take measures that would benefit them, and they do things that harm themselves. Such lapses derive from many causes, both cognitive and emotional, about which we have learned a lot in recent years from psychologists and behavioral economists. Often it seems these lapses could be remedied by simple measures.

* Department of Philosophy, Georgetown University. I am grateful to participants in the Harvard Political Theory Workshop and the Georgetown University Institute for the Study of Markets and Ethics symposium on nudging for helpful discussion of earlier drafts of this paper.

What, if anything, should be done to improve the chances that people act in their own best interests?[1] Three main answers to this question have been offered. Hard paternalists defend sometimes coercing people by law to do things (or not do them) for their own good. Soft or libertarian paternalists instead recommend "nudging" them—structuring their choices by manipulating the environments in which they choose, in ways they may or may not be aware of, so that they are more likely to make better choices. The third approach opposes forcing people or even nudging them to act; it only endorses providing them with information so they can choose wisely. The assumption underlying this approach is that people's poor choices arise only from ignorance rather than from other cognitive or emotional shortcomings.[2] Sarah Conly calls this view liberalism, no doubt in part because of its association with liberalism's most important defender, John Stuart Mill, who famously argued that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant."[3] But many hard and soft paternalists would also describe themselves as liberals. "Libertarianism" is a better description of the information-only approach—even though writers like Richard Thaler and Cass Sunstein have appropriated the term "libertarian paternalism" for their pro-nudging view—and I will use it here.[4]

In this paper I aim to make some progress in comparing the three approaches from a moral point of view, focusing primarily on libertarianism and soft paternalism. I begin with some reflections on the libertarian, information-only approach.

I. THE LIBERTARIAN VIEW

How do libertarians respond to what seems like the obvious fact that people do not always act in their own best interests? A few might bite the bullet and say, "Whatever you do is what you really desire to do, and there is no other criterion of acting in your own best interests apart from doing what you most desire to do." This response makes the proposition that people always act in their own best interests unfalsifiable, true by definition, and/or just plain implau-

1. I am not defending ethical egoism, the claim that people *ought* always to put their own interests ahead of others' interests. For my purposes we can limit ourselves to those situations in which it is legitimate and desirable for people to do what is best for themselves.

2. *See* DANIEL KAHNEMAN, THINKING FAST AND SLOW 411–13 (2011).

3. SARAH CONLY, AGAINST AUTONOMY: JUSTIFYING COERCIVE PATERNALISM 23–33 (2013); JOHN S. MILL, ON LIBERTY (Elizabeth Rapaport ed., 1978). It is doubtful, however, that Mill opposed all forms of paternalism. In the *Principles of Political Economy*, he offers several exceptions to the principle of laissez-faire, including: when the consumer is an incompetent judge of the commodity, especially respecting things that tend "to raise the character of human beings," such as education (947); where people exercise power over others; and contracts in perpetuity. *See id.* at Book V, Chapter 11.

4. *See, e.g.*, Cass Sunstein & Richard Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159 (2003); RICHARD THALER & CASS SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008).

sible. Intelligent libertarians respond to what I take to be the undeniable fact that people don't always act in their own best interests in two ways, which are not mutually exclusive.

One may be called deontological. On this approach, the freedom to act without external constraint has an ultimate value that trumps any bad consequences it may lead to. No matter how badly people mess up, human freedom should not be infringed.[5] Another deontological approach, associated with Robert Nozick, says that the state has no right to force people to act for their own good because doing so violates their rights.[6] (I shall not attempt to explain the relationship, if any, between these views.)

An alternative approach foregoes the absolutist argument and instead contends just that paternalistic interferences inevitably do more harm than the bad choices people make. This consequentialist line implicitly invokes a cost-benefit analysis and finds paternalism wanting. Several reasons why can be offered; we can find some of the principal ones in Mill's objections to paternalism in *On Liberty*.

First, the individual "is the person most interested in his own well-being."[7] In other words, you care more about your welfare than others do. Mill has forgotten the obvious counterexample: your mother![8] Still, we might concede that the claim is generally true: you care more about yourself than almost anyone else, certainly more than the politicians and policymakers who would be the likely paternalistic interveners.

Second, says Mill, "with respect to his own feelings and circumstances the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else."[9] Here we can't let Mill off so easily. Sure, there are reasons he thought what he did: Mill lived before Freud, not to mention Kahneman and Tversky and the rest. But even if we were to agree that people know their own feelings better than others do, their lack of knowledge of "circumstances" and many other things relevant to achieving their own good often prevents them from doing so. Self-deception, overconfidence, framing effects, loss aversion, the availability heuristic, and other reasons people fail to do what's good for themselves are likely to be familiar to anyone

5. Probably most of those who reject libertarianism also believe that freedom has intrinsic value and that it can trump bad consequences, but they deny that its value is absolute. This approach would put the thumb on the scale of freedom as an intrinsic good but allow that sufficiently bad consequences can outweigh it. People will disagree about how bad the consequences have to be to justify intervention.

6. ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA ix (1974).

7. MILL, *supra* note 3, at 74.

8. But then Mill altogether neglects his own mother, who is never mentioned in his *Autobiography*. As the editors of the definitive edition of Mill's works note, the Early Draft of the *Autobiography* contained a half a page about his mother but the passage was deleted in the final version, and despite his feminism Mill gives himself "but a single parent": "I was born, in London, on the 20th of May 1806, and was the eldest son of James Mill, the author of *The History of British India*." MILL, AUTOBIOGRAPHY xvii–xviii, 5 (1981).

9. MILL, *supra* note 3, at 74.

reading this paper.[10]

Mill's third argument against paternalism is that "All errors which he is likely to commit against advice and warning are far outweighed by the evil of allowing others to constrain him to what they deem his good."[11] This sounds much like what he says a few pages later is the "strongest of all the arguments against the interference of the public with purely personal conduct"—"that, when it does interfere, the odds are that it interferes wrongly and in the wrong place."[12] The talk of weighing suggests a consequentialist approach. But it is impossible to avoid the suspicion that Mill means something stronger, that he values freedom above all and without qualification. "The only part of the conduct of anyone for which he is amenable to society is that which concerns others. In the part which merely concerns himself, his independence is, *of right, absolute.*"[13] Mill's passion here and elsewhere suggests to many that despite his own protestations and the typical philosophical classifications, he is a closet non-utilitarian. (So much for self-knowledge.)

A consequentialist approach is ill-suited to the "absolute" judgment that paternalism is *never* justified. At the very least, it requires a rule-consequentialist view asserting that even though paternalism might occasionally be beneficial, it's hard to know exactly when and so in the long run a complete ban on paternalism works better. Although Mill may well be a rule-consequentialist (to the extent that he's a consequentialist at all), and although as a political philosophy rule-consequentialism is probably the only kind that works, a complete ban on paternalism still seems to me unconvincing.

II. HARD AND SOFT PATERNALISM: EFFECTIVENESS

At first blush, soft paternalism seems to have the advantage over hard paternalism. Coercion is harsh: surely it's better not to force people to do things, not to bring the heavy hand of government down on them, if we can induce them to act wisely without forcing them. But there are at least two reasons why nudging may seem problematic. First, it might be less effective or less efficient than coercion. That makes sense—again at first blush. Coercion seems to give people stronger incentives to act in the desired way. Second, many critics believe that nudging is manipulative or deceptive or in some other way nefarious, whereas coercion is out in the open.

Whether hard paternalism is more effective or efficient than soft is, of course, an empirical question. And it is likely there's no simple answer: it will depend on the circumstances and the particular kind of case. Consider seatbelt laws.[14]

---

10. An excellent summary of the literature can be found in KAHNEMAN, *supra* note 2; *see also* THALER & SUNSTEIN, *supra* note 4.

11. MILL, *supra* note 3, at 75.

12. *Id*. at 81.

13. MILL, *supra* note 3, at 9 (emphasis added).

14. The examples of seatbelt laws, motorcycle helmet laws, and many others often offered in discussions of paternalism are complicated by the fact that there are strong nonpaternalistic reasons for

Probably if nudging people to wear them were as effective as requiring them by law, that's what we would do. Insurance companies could lower rates for those who wear seatbelts, although verification would be a problem. Perhaps those found not to be wearing seatbelts when an accident happens would be penalized. Another way of nudging people is with slogans ("Seatbelts save lives") or videos (on TV ads or in driver's ed courses) showing the grisly consequences of not wearing seatbelts. The question, of course, is whether these approaches are as effective as coercion; it seems the answer is no.[15]

III. NUDGING VERSUS INFORMING

But this example raises an important question. Does sharing slogans or showing videos constitute true nudging, or is it simply informational, and thus legitimate even on the libertarian view? More generally, how can we distinguish nudging from informing? I will explore these questions through two examples.

First, consider an experiment in which people in remote Ethiopian villages were shown inspirational documentaries about others in nearby villages who had escaped poverty by setting goals, making crucial choices, and working hard. A placebo group was shown an entertainment video, and a control group was just surveyed. Returning to the village six months later, the researchers found that seeing the inspirational documentaries not only raised the villagers' aspirations but also had small but positive effects on savings and credit behavior and on enrollment of children in school and investment in their education.[16]

Were these documentaries simply informational? It might seem so: after all, they were factual, not fictional.[17] On the other hand, they were selective in the information they provided, and were chosen accordingly. The researchers might have shown villagers depressing documentaries as well—equally "true"—about efforts that had failed. Is that required under the information-only approach? Despite talk about "full information," in the real world, information is always partial and incomplete. The libertarian might reply that those supplying information must at least attempt to be "fair and balanced." It would seem to follow that the experimenters would indeed have to show depressing videos as well as

---

having such laws—that they reduce injuries and thus healthcare costs, for example, thereby benefitting many others besides those who fail to wear seatbelts or helmets. These collateral effects are so widespread it is difficult to identify purely paternalistic policies; it makes more sense to talk about paternalistic and nonpaternalistic *reasons* for policies rather than policies. *See* Gerald Dworkin, *Paternalism*, 56 MONIST 64, 64, 65 (1972). For this reason, it's not always easy to see whether or to what extent the paternalistic arguments are doing the work in the discussion.

15. *See* CONLY, *supra* note 3, at 149.

16. Stefan Dercon et al., *The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia* (Ctr. for the Study of Afr. Economies, Oxford Univ., Working Paper WPS/2014-16, 2014). I am grateful to Karla Hoff for making me aware of this study.

17. The experimenters ran a competition and on that basis had fifteen-minute documentaries made about ten individuals in other areas of the same region where the subjects of the experiments lived. Four documentaries were ultimately chosen (two about men, two about women) to be shown to the experimental subjects. *Id.* at 7.

uplifting ones. That seems a peculiar requirement. But if information is always partial and selective, then the distinction between informing and nudging may be difficult to draw.

A second example reinforcing this conclusion is brilliantly recounted in a 1991 article by legal scholar William Simon. (Simon's narrative is complex and subtle, and I do not do it full justice here.) Early in his career, Simon defended a woman in a criminal case who worked as a housekeeper for a partner in the law firm where Simon worked. Mrs. Jones, a "churchgoer" and "homeowner" in her sixties who had never had a brush with the police, "was charged with leaving the scene of a minor traffic accident without stopping to identify herself."[18] She claimed the other driver had caused the accident and had been the one to leave the scene. Mrs. Jones was black; the other driver was white. The question was whether Mrs. Jones should accept a plea deal. If she did, she "would have a criminal record, but because it would be a first offense, she could apply to have it sealed after a year."[19]

Mrs. Jones brought her minister to the courthouse, and they talked with Simon. They hoped he would tell her what to do, but he insisted it was her decision and he couldn't make it for her. Instead, he "spelled out the pros and cons" of accepting the plea deal—in that order, pros first and cons after. His final statement was: "If you took their offer, there probably wouldn't be any bad practical consequences, but it wouldn't be total justice."[20]

Mrs. Jones struck Simon as "a person who prized her dignity," and the statement made a big impression on her: she was ready to refuse the deal. Simon then brought over his friend, a more experienced lawyer with whom he had been working on the case, and told him her response. The friend "stared in disbelief" and then gave Mrs. Jones his advice. He mentioned the same considerations Simon had. But "he discussed the disadvantages of trial last, while I had gone over them first; he described the remote possibility of jail in slightly more detail than I had, and he didn't conclude by saying 'It wouldn't be total justice'."[21] Mrs. Jones decided to accept the plea bargain.

The most important lesson of Simon's story is that it is not only the content of information but also how it is presented that is crucial to how it influences those who receive it. The very same facts presented in a different order can affect people differently. Body language, vocal and facial expressions make a big difference as well. Simon's assertion about "total justice," especially at the end of his explanation and in light of his judgment of his client's character, clearly had an effect.

All this is pretty obvious. But it shows that the libertarian emphasis on an information-only approach is naïve. Wherever time and space are limited—that

18. William H. Simon, *Lawyer Advice and Client Autonomy: Mrs. Jones's Case*, 50 MD. L. REV. 214 (1991).

19. *Id.* at 215.

20. *Id.*

21. *Id.* at 216.

is, in all real-life settings—relevant information must be whittled down to a manageable quantity. Sometimes not all relevant information can be included, and information must be presented in some order or other and will almost always be inflected with features that are not purely informational but that affect people's understanding, desires, and choices. In the real world the distinction between informing people and nudging them almost always breaks down.

IV. NUDGING AND MANIPULATION

But the idea of nudging people makes some uneasy. As against either informing them or coercing them, nudging may seem somehow underhanded, deceptive, dishonest. It evokes talk of manipulation. And that's a criticism. My argument so far has been that the idea of neutrally disseminating information is often a myth, so if nudging is manipulative, informing may be as well. Hard paternalists might seem on firmer ground: they make no bones about forcing people to do things (or refrain from doing them), and that will generally be known to those forced. So insofar as nudging is not transparent to those nudged perhaps the charge of manipulation is well-founded.

What is manipulation? There is no consensus, although a significant philosophical literature now exists on the subject.[22] It has been variously argued that manipulation entails deception; that it essentially involves harm; that it undermines a person's autonomy; that it subverts a person's rational capacities; that it is necessarily intentional; and that it fails to track reasons.[23] It might seem necessary, if we are to evaluate nudging, to sort out these suggestions and determine the essential features of manipulation. But I think this is a mistake, for several reasons. First, I doubt that it's useful to think of manipulation in terms of necessary and sufficient conditions. The concept seems to fit better with a Wittgensteinian family resemblance view of meaning: some practices we are inclined to call manipulative may have features A, B, and C; others have B, C, and D but not A; others A, D, and E; etc.

Second, even if we settled on essential descriptive features of manipulation, the question would remain whether to accept a moralized or non-moralized conception. Is to say that Smith manipulated Jones to say that Smith did something bad or wrong or at least wrong-making even if ultimately justified? Here I am sympathetic to Marcia Baron's view that manipulating people is not invariably wrong (either by definition or otherwise), but it is generally a bad thing.[24]

---

22. *See, e.g.*, MANIPULATION: THEORY AND PRACTICE (Christian Coons & Michael Weber eds., 2014).

23. *See, e.g.*, Moti Gorin, *Towards a Theory of Interpersonal Manipulation*, *in* MANIPULATION: THEORY AND PRACTICE, *supra* note 19, at 73–97 (discussing these approaches and a defense of the last). Each of these probably has something to be said for it, except for the harm view—since by hypothesis paternalism, if successful, prevents a person from harming herself, we can rule out harm unless we make it true by definition that the very act of nudging is harmful.

24. Marcia Baron, *The Mens Rea and Moral Status of Manipulation*, *in* MANIPULATION: THEORY AND PRACTICE, *supra* note 19, at 107–08.

If we cast our question about whether nudging is legitimate in terms of manipulation, we would first have to offer a theory of manipulation and then see whether nudging exemplifies it. Since the meaning and moral status of manipulation is itself under dispute, I see nothing to be gained by this approach. Many different kinds of practices and actions come under the heading of nudging, so a general theory of manipulation will not be very helpful.[25] Is a given practice in fact manipulative (does it have those essential features)? If so, is that necessarily bad (should manipulation be understood as an essentially moralized concept?)? Even if manipulation is essentially bad, is it necessarily so bad as to make the practice in question morally impermissible? And how should we judge the badness of manipulation as compared with the badness of coercion?

For these reasons, instead of engaging in an analysis of manipulation, I will consider some of the main kinds of nudging practices that have been advocated and see whether they have properties that ought to trouble us (and, if so, how they compare with alternative approaches, specifically coercion).

V. THE INEVITABILITY OF NUDGING

The first point connects to my earlier discussion comparing informing and nudging. Just as there is rarely a completely neutral way to present information, the environments in which people make choices of all kinds inevitably create their own nudges.[26] Consider the well-known example of the cafeteria line that begins Thaler and Sunstein's book *Nudge*.[27] Carolyn, the director of food services for a city school system, must decide how to arrange food in the cafeteria line. Assume that how food is arranged affects what and how much people eat: if you put healthy foods like fruits and vegetables at the front of the line and desserts at the end, children will choose and eat healthier food. Is it wrong for Carolyn to arrange the food in this way? Is doing so an unjust or intrusive use of her power?

What's the alternative? As Thaler and Sunstein point out, Carolyn could order the food differently. She could aim to maximize profits; she could choose the order at random; she could arrange the food in a way that would get students to choose the foods they would choose on their own. It seems clear that the first option involves nudges of its own. The second seems designed to show, with a

25. Sidney Morgenbesser, *Scientific Explanation*, *in* 14 INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES 122, 122 (David Sills ed., 1968) ("To explain why a man slips on a banana peel we do not need a general theory of slipping.").

26. Some will dispute this, arguing that nudges are necessarily intentional, so that features of the environment, or indeed unintentional acts of agents, that influence people to act in one way rather than another are not nudges, whatever else we may say about them. I hope to show in what follows that the dispute is largely terminological and that in many cases there is no good reason to call only the intentional acts of agents nudges. But the defender of nudging could just as easily dispense with the term; instead of asserting that nudging is inevitable she could adopt John Hasnas's formulation that "some default is inevitable." *See, e.g.*, John Hasnas, *Some Noodging About Nudging: Four Questions About Libertarian Paternalism*, 14 GEO. J.L. & PUB. POL'Y 645 (2016) (exploring this issue).

27. THALER & SUNSTEIN, *supra* note 4.

vengeance, that Carolyn aims to avoid influencing students in any way ("It's not for me to say what they should eat"), but it hardly seems a satisfying approach. And randomness could produce its own nudges, although they would be unintended, unlike the others. The third approach might in fact seem best to promote kids' autonomy and avoid nudging them.

The example might seem ill-chosen, insofar as most people think paternalism is justified with respect to children. (The objection to paternalism may be put by saying that it treats people *as if* they were children.) So suppose we make Carolyn the director of food services in a cafeteria for government workers. Would it then be appropriate for her to arrange the food in a way that gets workers to choose the foods they would choose on their own?

There are at least two reasons to think not. First, as Thaler and Sunstein note, to the extent that people's choices depend on the order in which food is displayed, the idea of mimicking their true choices is misguided. This is a concrete example of a much more general point: what people want is partly shaped by what is presented to them, what choices they are offered.[28] So the idea that there are always true preferences to be discovered is illusory. Daniel Hausman and Brynn Welch argue that placing placards with nutritional information next to food items would be preferable to arranging the items in order of their health benefits, and that doing so would not be paternalistic.[29] But this suggestion neglects the fact that the food must be arranged in some order or other, and that its order may create its own nudges, whether intentional or not.

"What matters," Hausman and Welch say, "is whether the policy-maker is attempting to bring about something against the beneficiary's will."[30] This claim brings us to the other central question about arranging food to get people to eat more healthily: *is* it against the beneficiary's will? Hausman and Welch seem confident that it is, assuming that the person on the cafeteria line wants to eat fries while Carolyn tries to get her to eat salad instead. They seem to assume people have consistent and unified desires. But this is an illusion. Our desires are often conflicting and inconsistent.

Let me eat cake. Let me live to be fit and healthy well into old age. I want both these things. This point is commonly put in terms of a distinction between short-term and long-term desires, paralleling immediate versus delayed gratification. Philosophers distinguish between first-order desires and second-order or higher-order desires: I may have a desire to smoke but I also have a desire to desire not to smoke.[31] Or the distinction may be drawn in terms of desires that one endorses and those one doesn't. The desires to eat cake and to live a long and healthy life don't quite fit these models, however; my desire to live a

28. *See, e.g.*, JON ELSTER, SOUR GRAPES: STUDIES IN THE SUBVERSION OF RATIONALITY (1985).

29. Daniel M. Hausman & Brynn Welch, *Debate: To Nudge or Not to Nudge*, 18 J. POL. PHIL. 123, 129 (2010).

30. *Id*. at 130.

31. *See, e.g.,* Harry Frankfurt, *Freedom of the Will and the Concept of a Person*, *in* THE IMPORTANCE OF WHAT WE CARE ABOUT (1988).

healthy life is a first-order desire, like the desire to eat cake. But the two desires range over different sets of particulars; my desire to be healthy is a long-term desire that is not incompatible with the desire to eat cake on every occasion when that possibility presents itself. Sometimes it's okay to eat cake! Still, as I move along the cafeteria line on any given day I may experience both desires, and they may conflict.

Perhaps we should ask about the preferences concerning food arrangement of those who frequent the cafeteria. Do they oppose putting the healthy foods first? If asked many would probably agree that this is a good ordering and appreciate having it in place. If so, that should be enough to satisfy Hausman and Welch: the ordering would not be against the agents' wills. But not everyone will like the ordering. Should we take a vote and let the majority prevail? It will not always be practicable or possible to decide such matters democratically. Does it really make sense to demand that Carolyn leave the decision about how to arrange the food to those who regularly frequent the cafeteria? If not, Carolyn (and others in charge) will have to decide on their own. So if they want to act in a way that isn't against their customers' wills, they will have to figure out what it is that the customers will. Of course the customers will not all share the same goals, but it does not seem unreasonable to suppose that most people want to live a long and healthy life.

This line of reasoning may seem too quick and too slick. It seems just another few steps to the idea that people don't always know what they want, to "real will" theories that drive a wedge between desires that people know they have and things they "really" want. Just short of real will theories are perfectionist theories, which "recommend the pursuit of lives of objective value rather than simply the satisfaction of desire."[32] I do not have the space here to enter into the debate between subjective and objective conceptions of well-being. I'll simply assert that even a subjective theory of well-being can perfectly well account for the legitimacy of paternalism, hard or soft. Nudging people to eat their broccoli will correspond to (some of) some people's central desires. If this conclusion still seems too easy, that's because philosophers may be reluctant to admit that most of us find within ourselves bundles of desires that conflict or at least stand in tension with each other.

The upshot of this discussion is that in many contexts nudges are inevitable features of the "choice architecture" in which we find ourselves.[33] We make choices inside environments that have their own characteristics and properties; these, in conjunction with our biases, limitations, and predispositions, incline us to choose one way rather than the other. To insist that nudges produced by

32. *See* Conly, *supra* note 4, at 103, which provides a good discussion and defense of a non-perfectionist approach. She sums up "the problem with objective standards of welfare" this way: it would be strange if "my life can be going well despite my failure to have any positive attitude toward it" (quoting L.W. Sumner, *The Subjectivity of Welfare*, 105 ETHICS 764 (1995)).

33. The term is from THALER & SUNSTEIN, NUDGE, *supra* note 3.

human will are illegitimate is to allow nature to take its bossy and often undesirable course. I return to this point below.

VI. DEFAULTS

Another prominent set of examples that illustrates the legitimacy of nudging concerns defaults. Default options exist, among other places, for mortgage rules, pension plans, and organ donation policies. (The latter are not paternalistic policies, because they concern acting for someone else's good, not one's own, but the same principles apply.) For example, mortgage rules can be designed with opt-out defaults that nudge people to choose a standard product and make it harder for borrowers to choose one they may not understand.[34] Retirement savings plans can automatically enroll employees rather than enrolling them only if they sign up. In one plan studied, "the percentage of employees saving for retirement increased from 49% to 86% when the default was changed to automatically enrolling employees."[35] Organ donation policies provide a dramatic example of the power of defaults. In some countries, including the United States and Great Britain, people must choose, when they get or renew their driver's licenses, to become organ donors; the default is not to donate. In many European countries, the policy is the reverse: consent to donate is presumed, and you must explicitly opt out to avoid donation. In Austria, France, Hungary, Poland, and Portugal, which all have opt-out policies, effective consent rates are over 99%. In countries with opt-in policies, consent rates are radically lower—from 4.25% in Denmark to 27.5% in the Netherlands.[36]

Although defaults nudge people rather than coercing them, some critics nevertheless find them overly intrusive or manipulative. They prefer giving people information about the consequences of choosing one way or the other and leaving it to them to decide what to do. We saw earlier that it's often difficult or even impossible to present information neutrally and nudge-free. Still, nudges can sometimes be avoided through mandated or forced choices. People may be required to complete a form that asks them explicitly to donate their organs or not, or to enroll in a pension plan or not. If filling out the form online, an applicant might not be able to proceed without providing an answer. On paper, however, some people might fail to check a box, in which case the motor vehicle bureau or employer would need a default option.

What's a good default? Perhaps it's the one I would prefer if I had full information and sufficient time and mental resources to process it. On this criterion, since people have different values and preferences, no default is

34. Michael S. Barr, Sendhil Mullainathan, & Eldar Shafir, *Behaviorally Informed Regulation, in* THE BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 449 (2013) (offering many other examples). Much of this section derives from Judith Lichtenberg, *Paternalism, Manipulation, Freedom, and the Good, in* BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 694–98 (2013).

35. *See* Shlomo Benartzi, Ehud Peleg, & Richard Thaler, *Choice Architecture and Retirement Savings Plans, in* BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 245 (2013).

36. E.J. Johnson & D.G. Goldstein, *Do Defaults Save Lives?* 302 SCI. 1338 (2003).

necessarily best for everyone. In an online experiment Johnson and Goldstein show that in organ donation mandated choice approximates the opt-out default: 79% of participants who must decide choose to be organ donors; 82% in the opt-out default remain as donors; only 42% in the opt-in condition agree to be donors.[37]

Mandated choice is not always possible, but even when it is, defaults may be preferable. Conly expresses a view that is certainly widely shared: "I hate having to review all my health insurance options, pension options, mortgage options, and credit card options" and "If the government were to do the research and ascertain that trans-fats are bad for my health and then remove trans-fats from my diet options, I'd be grateful."[38] If more people agree than not, perhaps that justifies defaults that nudge.

Of course, Conly here implies that the government is trustworthy, which many will with good reason deny. But the problem is a much broader one that affects everything we need to know and claim to know. It's not simply that most of us would rather spend our days doing something other than puzzling through multiple pension plans; given constraints of time and talents, almost no one is capable of making good judgments without relying on others' expertise. Even in the libertarian's ideal nudge-free just-the-facts environment, few are able to evaluate all the facts needed to form judgments and make decisions; instead we have to put our trust in others. That's the nature of modern life.

VII. PERSUASION, RATIONAL AND OTHERWISE

Lurking in the discussion so far is a deeper question that I can make only small inroads into, at best. It's raised by Hausman and Welch, who assert that rational persuasion is the ideal way for government to influence the behavior of citizens. Although the force of rational persuasion is limited, and actual persuasion is rarely purely rational, only rational persuasion fully respects the sovereignty of the individual over his or her own choices.[39]

Philosophers are especially susceptible to this view, and of course it has held sway in economics at least until the work of Kahneman and Tversky and the behavioral economics wave began to topple its hegemony. To defend nudging is to concede that, humans and circumstances being what they are, we cannot rely only on rational persuasion and the power of the best argument to convince people that they should do *x* rather than *y*. We have seen several good reasons for this conclusion, even apart from human shortcomings and biases, including that the environments in which people make choices inevitably contain their own nonrational nudges, whether intentional or not, and that people often have neither the time, inclination, or expertise to consider all the evidence in a way

37. *Id.*
38. CONLY, *supra* note 3, at 90–91.
39. Hausman & Welch, *supra* note 26, at 135.

that would be necessary to reach rational decisions. The question still remains whether rational persuasion is the ideal to which we should aspire.

A story from my own experience illustrates the dilemmas. For about as long as I have been teaching, I have taught the death penalty in my ethics and philosophy of law courses. I find that it illustrates certain fundamental moral problems and concepts (such as deontology and consequentialism) as well as any topic I know, and it is real and riveting to students. I oppose the death penalty—for just about every reason you can give for opposing it—but I take care to present the issues fairly, and I don't tell students my view. I present the arguments, the objections, the replies to the objections, not ad infinitum, but as thoroughly as I can in the classroom setting.

In the 1980s, a faculty member at the University of Maryland Law School who specialized in death penalty appeals guest-taught in several of my undergraduate courses. Except for some time at the end of class, where he queried students and took questions, Mike Millemann spent the whole time describing a single case he had litigated, a death penalty appeal in Florida. Ernest Fitzpatrick was a mentally-ill twenty-year-old who conceived the idea of taking hostages in a real-estate office to get money. The plan blew up, and a sheriff's deputy was killed. The jury convicted Fitzpatrick of first-degree murder and sentenced him to death. Millemann, asked to help by a death-penalty activist and Fitzpatrick himself in a handwritten note, donated over 2000 hours to the case. In 1988 the Florida Supreme Court reduced Fitzpatrick's sentence to life with the possibility of parole.[40]

When he spoke to my class, Millemann would simply tell the story of his involvement in the case. No editorializing, no moralizing; he would just describe the course of events. The students were spellbound, and at the end of the hour many rushed up to talk to him, wanting to know how they too could become death penalty appeals lawyers. They seemed miraculously converted. By contrast, even though I thought the arguments against the death penalty spoke for themselves (despite objections and arguments on the other side) I never had the sense that my rationalist approach to the subject moved my students much to change their minds. Of course, I might have been wrong: it may have been that the changes weren't obvious to me, or that it took time for the arguments to sink in and that some students did change their minds.[41]

---

40. For an account of the case, see Jonathan Pitts, *Legal Activist Mike Millemann Has Spent His Career Striving to Provide 'Access to Justice*,' BALTIMORE SUN (July 28, 2015), http://www.baltimoresun.com/news/maryland/politics/bs-md-millemann-20150725-story.html. As I recall, Mike visited my class before the Florida Supreme Court decision was handed down, so we did not even know at the time that his appeal succeeded.

41. I do not mean to suggest that everyone who listens carefully to the arguments will be converted to the anti-death penalty view or that there's nothing to be said on the other side. Still, since many students came in with naïve views of the matter I think it's reasonable to expect some changes. I should add that in recent years it seems to me that, perhaps because of growing media coverage of the criminal justice system and common public criticisms of it, college students begin with a view that's more critical of the death penalty than they did back in the 1980s.

I found the experience somewhat depressing. What's the point of carefully dissecting arguments if that's not what affects people's thinking? Of course it fits with much of the behavioral economics and psychology literature showing that people are moved to act much more by individual stories and pictures, and by all sorts of other subtle cues, than by statistics and arguments.[42] And it confirms Richard Rorty's view that changes in moral outlook arise from seeing pictures or hearing stories, not from considering or accepting arguments.[43]

It's interesting that even Hausman and Welch, who treat rational persuasion as the gold standard, pull back: "We do not mean to suggest that rational persuasion is emotionless cold calculation. Clarifying the role of emotions in rational persuasion is a difficult task for another occasion."[44] Yes, it is a very tall order, and I too can only say a little here. (And it's not just the role of emotions we need to understand, but the role of all factors that are not properly considered "reasons.") We might begin with the idea that some emotions are appropriate to certain situations and others are not. So it's appropriate to draw someone's attention to the lived experience of being under penalty of death, as Millemann did in my class. On the other hand, a full and fair treatment of the issue would have to consider the experience of murder victims and their families too. One might ask whether, similarly, when attempting to persuade people to give to alleviate global poverty we have, in fairness, to show not only pictures of children like Rokia but also snapshots of the consequences of other uses of the funds that might have gone to her. Before spending hundreds of thousands of dollars rescuing a toddler who has fallen into a well, we should tell the public that we could save many more children by using those funds to reduce lead-paint poisoning.[45]

It's possible that these treatments would cancel each other out: we will be emotionally affected by thoughts of death row, but also by thinking of the families of murder victims. So it may not help to say that the stories we tell or the pictures we show have to be appropriate and fitting, because appropriate and fitting stories and pictures can always be found for various "sides" of a story.

---

42. *See, e.g.*, Deborah A. Small, George Loewenstein & Paul Slovic, *Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims*, 102 ORG'L BEHAVIOR & HUM. DECISION PROCESSES 143, 143–53 (2007). Experimental subjects were given the chance to give $5 to the organization Save the Children. In one treatment, subjects saw only an identifiable victim, a girl named Rokia; in a second, they saw statistics about victims of hunger and malnutrition; in a third, they saw both. Not surprisingly, donations were far greater among subjects in the first treatment than among those in the second. But they were also greater among those in the first treatment than among those in the third; in other words, seeing statistics *in addition to* an identifiable victim reduced people's likelihood to contribute.

43. Richard Rorty, *Human Rights, Rationality, and Sentimentality*, *in* ON HUMAN RIGHTS: THE OXFORD AMNESTY LECTURES (Stephen Shute & Susan Hurley eds., 1993).

44. *See* Hausman & Welch, *supra* note 26, at 135.

45. *See* J. Michael Kennedy, *Jessica Makes It to Safety—After 58 ½ Hours*, L.A. TIMES (Oct. 17, 1987), http://articles.latimes.com/1987-10-17/news/mn-3702_1_jessica-mcclure. Those who were conscious in 1987 will remember Jessica McClure, the eighteen-month-old who fell into a well in Texas and was rescued at great expense after fifty-eight hours and viral media attention.

These sorts of considerations may leave one pessimistic about the possibility of "clarifying the role of emotions in rational persuasion," as Hausman and Welch call for.

VIII. "IMPOSITION OF THE WILL"

I have been arguing that nudging is inevitable and is therefore permissible, because "ought" implies "can" (or, more precisely in this case, "ought not" implies "cannot"). But this view implies that there is no morally relevant difference between those influences that arise out of our cognitive shortcomings, or the choice architectures or environments within which we make our decisions, and those imposed intentionally by human agents.

But some object to nudging precisely on the grounds that these are not equivalent; they deny that effects of cognitive shortcomings or biases or choice architecture are properly called nudges. For example, Hausman and Welch argue that there's "an important difference between choices that are intentionally shaped and choices that are not"—the former "imposes the will of one agent on another" and is therefore morally problematic in a way that choices not intentionally shaped are not.[46] If true, this conclusion would essentially show that nudging is not inevitable, because according to these critics only intentional nudging counts as nudging in the relevant sense.[47]

In response, I shall offer three lines of argument:

1. The first is to question whether "imposition of the will" is an accurate description of nudging.
2. The second is to acknowledge that this kind of influence is a bad-making feature of an act but deny that all such acts are therefore unjustified.
3. The third is to question whether this kind of influence is necessarily problematic (bad-making).[48]

In this section I develop the first, in the following sections the second and third responses.

Does nudging impose the will of one agent on another? We are not talking here about coercion, but about influence short of it. Imposing one's will

46. *See* Hausman & Welch, *supra* note 26, at 133. I am grateful to Victor Tadros for making this point forcefully in discussion.

47. Similarly, many argue that to count as manipulation an act must be intentional. *See, e.g.*, Baron, *in* MANIPULATION: THEORY AND PRACTICE, *supra* note 21, at 98. For the view that one can manipulate without intending to and even while intending not to, see Kate Manne, *Non-Machiavellian Manipulation and the Opacity of Motive*, *in id.* at 221–46.

48. In discussing manipulation, Baron similarly distinguishes between these two interpretations—that manipulation is "always objectionable and to be avoided, but sometimes the best option," and that it is "more often than not objectionable but sometimes not at all objectionable." Baron, *in* MANIPULATION: THEORY AND PRACTICE *supra* note 21, at 108. She argues that there are examples of both kinds of cases. Although manipulation and nudging are not identical, I adapt her distinction here.

suggests that the nudgee cannot escape the power of the nudger. But nudging is perhaps better described as "attempting to influence another person by other than rational means." The question then arises whether it is so clear that doing so is morally problematic. Of course, we do not want to (and cannot) decide whether nudging is problematic by terminological fiat. But it's fair to say that describing nudging as imposition of the will is at the least tendentious.

IX. THE LESSER OF EVILS

But suppose we accept that description for the moment. Few people think that one agent's imposing its will on another is always morally impermissible. If so, law and government would be impermissible. Only philosophical anarchists believe that, and I assume most participants in these debates are not anarchists. The critic may respond in Millian fashion: "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others." That, of course, is precisely the question at issue: whether the person's own good is or is not "a sufficient warrant."

What reason do we have for thinking that preventing harm to others may outweigh or override the badness of imposing one's will on a person, but that promoting that person's own good never can? Imposition on the will of another certainly requires justification, but the question is how high the bar should be to justify it and whether there is a difference between doing it for others' sake and doing it for the agent's sake that is sufficient to allow the first but never the second.

It might be argued that (at least some) actions that harm others come under the heading of justice, and that these are by their very nature the category of actions that justify coercion, whereas self-regarding actions do not.[49] Yet even if harmful-to-others actions, or matters of justice, constitute a sufficient condition for imposing constraints on others, it doesn't follow that these constraints are necessary, so that we are warranted in imposing A's will on B for C's and D's and E's sakes but never for B's sake. That is precisely the question at issue in nudging that "imposition of the will" critics seem to beg.

On the assumption that nudging is a bad-making feature of actions, its defenders will argue that its benefits sometimes outweigh its costs. If we can induce people to eat more healthily by arranging the food in the cafeteria one way rather than another; if we can get people to save more for retirement by capitalizing on status quo bias or loss aversion—and these do not involve deception or other nefarious practices—what reason do we have to think that the bad of nonrational influence (which we are here assuming for the sake of argument) can never be outweighed by its benefits? I doubt that most of nudging's critics are committed to such a strong view.

49. Laura Valentini made this point in discussion.

We should remember also that in the cases under scrutiny any "imposition of the will" or nonrational influence is by hypothesis exerted for the benefit of the person influenced. In this it is unlike typical instances of imposition of an agent's will (and of the associations with it), which are usually designed for the benefit of the agent. (Thus note Hausman and Welch's statement, quoted earlier, that "What matters is whether the policy-maker is attempting to bring about something against the beneficiary's will."[50])

Suppose for the moment we grant the critics' claim that there is an important moral difference between the influences exerted by choice environments and human shortcomings, on the one hand, and intentional nudges on the other. Still, the dichotomy is highly misleading. It's not as if, without nudging from governments or other public bodies, the only remaining influences are "natural" ones. PepsiCo and Hostess Brands will impose their wills to put twenty-ounce cans of Mountain Dew and double packs of Twinkies in vending machines front and center in the cafeteria.[51] The goal of these companies, needless to say, is not to advance the good of those who consume their products. So in refusing to engage in nudging we may be leaving it to others whose motives do not include advancing the welfare of the nudged.

Summing up the points made so far in this section: Even if we think exerting non-rational influence of the sort at issue in nudging is a bad-making feature of actions, it's extremely implausible to think it renders all such actions impermissible. To do so, critics would need to show that such exertion of influence is so evil as to outweigh any bad consequences of not nudging, or that there is a principled and decisive difference between paternalistic nudging and nudging to prevent harm to others that renders the latter but not the former permissible. And they would have to say that paternalistic nudging is impermissible even to prevent or counteract the nudging by corporate or other bodies with their own interests at heart.

X. IS NUDGING ALWAYS INTRINSICALLY BAD?

Are there circumstances in which paternalistic nudging is not a bad-making feature of actions at all but is morally neutral and thus not in need of justification? Baron suggests that some instances of manipulation are like this (and it would seem nudging is less fraught with negative connotations than manipulation so that the case could be made more easily). She gives the example of a real estate agent who boils water with a drop of cinnamon or vanilla in it to infuse a house with a pleasant aroma while showing it to potential buyers (and where the agent is not trying to mask a bad odor in the carpets, for example).[52] Perhaps the question is not that important for our purposes; it's enough to show that nudging is justified even if in the ideal world it would be unnecessary. Still, I

50. Hausman, *supra* note 26, at 130.
51. I am grateful to Laurie Shrage for emphasizing this point in discussion.
52. Baron, *in* MANIPULATION: THEORY AND PRACTICE, *supra* note 21, at 114.

think a case can be made that nudging in some cases is not simply the lesser of evils. I shall explore two examples, the second of which I discussed earlier in a slightly different context.

Consider first an example recently reported in the *New York Times*.[53] In a three-month experiment, some vendors who provide the U.S. government with goods and services received a slightly different form than the usual one to report rebates they owed the government. The only difference was that the signature box was at the beginning rather than the end of the form. The result, according to reporter Binjamin Appelbaum: "a rash of honesty. Companies using the new form acknowledged they owed an extra $1.59 million in rebates during the three-month experiment, apparently because promising to be truthful at the outset actually caused them to answer more truthfully."[54] It's unclear whether the technique works in the long term; moreover, its purpose is not paternalistic (except in the extended sense that "we" all benefit if citizens pay what they owe to the government). But the relevant question here is whether there is *anything* problematic with requiring signatures at the beginning of a form rather than the end if it promotes more honest reporting. A similar question arose in discussing defaults, where we saw that in some cases default options are unavoidable.[55]

Is it really plausible to think that in cases like this the people must express their collective will to endorse the location of the signature on the form in order to render it legitimate? No. This is an example of the ubiquity of choice architecture—we fill out tons of forms, we make choices in myriad environments real and virtual, and these environments have design features that may influence us in significant ways. These features are not simply part of nature (the allegedly acceptable alternative to "imposition of the will"); they must be constructed and designed, whether advertently or not, intelligently or not. It does not seem plausible that designing them to promote uncontroversial benefits like honesty is even slightly problematic.

The other kind of case where we might want to say there is nothing morally problematic about nudging has to do with the role of emotion in reason, a subject discussed earlier in connection with Hausman and Welch's critique of nudging. Recall that despite their hostility to nudging and their view that rational persuasion is the gold standard, Hausman and Welch acknowledge that emotions have a place in rational persuasion.[56] We saw that it's extremely difficult to specify the appropriate role of emotion in rational persuasion, and I have little to add at this point. But if there is an appropriate role, it seems to follow that nudging is sometimes not even prima facie wrong.

---

53. Binyamin Appelbaum, *Behaviorists Show the U.S. How to Improve Government Operations*, N.Y. TIMES (Sept. 29, 2015), http://www.nytimes.com/2015/09/30/business/behaviorists-show-the-us-how-to-improve-government-operations.html.

54. *Id.*

55. As we saw in the cases of organ donation and retirement pensions, it is possible to require that a person make a choice, but defaults are sometimes inevitable.

56. Hausman & Welch, *supra* note 26, at 135.

XI. NUDGING VERSUS COERCING

Nudging is legitimate because it's inevitable, and because "ought not" implies "can not." Certainly in the real world, and perhaps even beyond it, there is often no way to present information neutrally and nudge-free. You often cannot impart all the information relevant to a decision, and even the order in which you impart information and how it is inflected can influence decisionmakers. And decisionmakers are subject to many pitfalls and biases that can lead them astray.

I have said little in this essay about how to decide between nudging and coercing, and I will end with only a few words on the subject. If my view that nudging need not be manipulative or nefarious is right, then I think it's safe to say that coercion is a more serious intrusion on liberty than nudging. So between the two there should be a presumption in favoring of nudging. Sarah Conly makes a forceful defense of hard paternalism, arguing at great length that its benefits in some cases easily outweigh its costs. Interestingly, her most prominent examples involve coercion of third parties (trans-fat bans, portion-size regulation, banning the sale of cigarettes), which seem somehow less intrusive and less objectionable than coercion that directly restricts agents who want to consume these products. Although Conly advocates a sophisticated cost-benefit analysis to determine when coercion is justified (taking into account, for example, the fact that people *just don't like* being told what to do or not do), and this will rule out coercion in many cases we care about, it's also true that she values autonomy less than most people, asserting that "preserving our liberty of action is not worth the costs of exercising choice."[57] Her view that "autonomy is not all that valuable" will seem shocking to many.[58] But even those who value autonomy more than Conly does can allow that nudging is often appropriate, either because the goals it serves override the value of autonomy or because, as she herself argues, nudging sometimes promotes autonomy itself.

57. CONLY, *supra* note 3, at 16.
58. *Id*. at 1.