

HEINONLINE

Citation:

Cass R. Sustein, The Ethics of Nudging, 32 Yale J. on
Reg. 413, 450 (2015)

Content downloaded/printed from [HeinOnline](http://heinonline.org)

Wed Oct 11 08:04:01 2017

- Your use of this HeinOnline PDF indicates your acceptance of HeinOnline's Terms and Conditions of the license agreement available at <http://heinonline.org/HOL/License>
- The search text of this PDF is generated from uncorrected OCR text.
- To obtain permission to use this article beyond the scope of your HeinOnline license, please use:

[Copyright Information](#)



Use QR Code reader to send PDF to
your smartphone or tablet device

The Ethics of Nudging

Cass R. Sunstein†

All over the world, governments are using nudges as regulatory tools. Is this ethical? Much of the answer depends on whether nudges promote or instead undermine welfare, autonomy, and dignity. Many nudges, and those that deserve support, promote some or all of those ideals, and undermine none of them. If welfare is our guide, much nudging is actually required on ethical grounds, even if it comes from government. If autonomy is our guide, much nudging is also required on ethical grounds, in part because some nudges actually promote autonomy, in part because some nudges enable people to devote their limited time and attention to their most important concerns. Finally, nudges should not, and need not, compromise individual dignity, which many nudges actually promote. There is, however, a genuine risk that some nudges might count as manipulation; an emphasis on welfare, autonomy, and dignity helps to show how to avoid that risk.

- I. The Central Argument..... 414
- II. Concepts and Definitions 417
 - A. In General..... 417
 - B. The Inevitability of Choice Architecture..... 420
 - C. Spontaneous Orders and Visible Hands 422
- III. The Trap of Abstraction..... 424
 - A. Nudges and Navigability..... 424
 - B. Three Distinctions..... 426
 - C. Illicit Reasons and Transparency 428
- IV. “As Judged By Themselves” 429
 - A. The Basic Standard..... 429
 - B. Questions and Doubts 430
 - 1. Objectively Good Lives? 430
 - 2. Ex Ante or Ex Post?..... 430
 - 3. Preferences About Preferences 431
 - 4. Informed Judgments and Empirical Puzzles..... 432
 - 5. Self-Control 432
- V. Applied Ethics..... 433
 - A. Welfare..... 433
 - B. Autonomy 437
 - C. Dignity 439
 - D. Manipulation 442
 - E. Biased Officials..... 449

| | |
|------------------|-----|
| Conclusion | 450 |
|------------------|-----|

I. The Central Argument

The goal of this Article is to explore the principal ethical objections to the use of nudges and choice architecture as regulatory tools.¹ As we shall see, the ethical issues largely turn on whether nudges promote or instead undermine welfare, autonomy, and dignity.² As we shall also see, the ethical analysis of nudges is similar to the corresponding analysis for other tools, such as fines and mandates, but there are some distinctive wrinkles, involving potential threats to autonomy and dignity. In particular, a concern for personal agency often motivates the most plausible objections to nudges.

The last decade has seen a remarkably rapid growth of interest in choice-preserving, low-cost regulatory tools.³ Especially in light of that interest, it is important to obtain an understanding of the nature and weight of the ethical objections to nudging. Many regulatory tools, and perhaps increasingly many, involve nudges and choice architecture, and some of the ethical objections can have either political or legal resonance. For example, nudges that involve information disclosure, or compulsory warnings, might raise First Amendment issues, triggering a concern with individual autonomy.⁴

† Robert Walmsley University Professor, Harvard University. This essay was the basis for remarks on the ethics of choice architecture and nudging, delivered at a conference on that topic at Humboldt University in Berlin in January 2015. I am most grateful to participants in the conference for many valuable thoughts and suggestions. I am also most grateful to Matthew Lipka, Martha Nussbaum, Lucia Reisch, and Adrian Vermeule for superb comments on a previous draft. I am also grateful to Heidi Liu for truly excellent research assistance.

1. Some of the strongest objections can be found in RICCARDO REBONATO, *TAKING LIBERTIES: A CRITICAL EXAMINATION OF LIBERTARIAN PATERNALISM* (2012).

2. Self-government might well be counted as an independent concern, though it might also be taken as derivative of one or more of the values of welfare, autonomy, and dignity. I do not give it separate treatment here, though as the examples suggest, we could imagine nudges that either promote or undermine the interest in self-government. For an interesting example, see the account of automatic voter registration in Oregon in Sheila V. Kumar, *Oregon is the First State to Adopt Automatic Voter Registration*, ASSOC. PRESS (Mar. 17, 2015), <http://www.bostonglobe.com/news/nation/2015/03/16/oregon-first-state-adopt-automatic-voter-registration/blEluzAXL7JzloPBIDh40I/story.html>.

3. Catalogues can be found in PETE LUNN, OECD, *REGULATORY POLICY AND BEHAVIORAL ECONOMICS* 25-38 (2014); DAVID HALPERN, *INSIDE THE NUDGE UNIT* (2015); Cass R. Sunstein, *The Council of Psychological Advisers*, ANN. REV. PSYCHOL. (forthcoming 2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2496438, at pp. 2-5. In 2015, President Obama issued an important executive order, directing his agencies to incorporate behavioral science. Exec. Order No. 13,707, 80 Fed. Reg. 56,365 (Sept. 15, 2015). At the same time, the White House's Social and Behavioral Sciences Team issued its annual report, offering a catalogue of recent interventions, all of them nudges. SOC. & BEHAVIORAL SCI. TEAM, EXEC. OFFICE OF THE PRESIDENT, *ANNUAL REPORT* (2015) [hereinafter *ANNUAL REPORT*].

4. On the FDA's effort to require graphic warnings on packages, see R.J. Reynolds Tobacco Co. v. U.S. Food & Drug Admin., 823 F. Supp. 2d 36 (D.D.C. 2011), *aff'd*, 696 F.3d 1205 (D.C. Cir. 2012). For the government's own graphic campaign, see CDC, *Tips from Former Smokers*, <http://www.cdc.gov/tobacco/campaign/tips/resources/videos/>. For evidence of success from graphic warnings, see Press Release, CDC, *Report Finds Global Smokers Consider Quitting Due to Graphic*

My central argument is that, at least if they are taken in general or in the abstract, the ethical objections lack much force, and for two different reasons. *First*, both nudges and choice architecture are inevitable, and it is therefore pointless to wish them away. *Second*, many nudges, and many forms of choice architecture, are defensible and even required on ethical grounds, whether we care about welfare, autonomy, dignity, or some other value.

It is true that all government action, including nudges, should face a burden of justification (and sometimes a heavy burden). If the government requires disclosure of information, or establishes particular default rules, it must explain and defend itself. The fact that people retain freedom of choice, and are ultimately permitted to go their own way, is important, but it does not give public officials a license to do whatever they want.⁵ In many cases, however, the requisite explanation is available.

Suppose, for example, that we believe that the goal of social ordering (including those forms for which government is responsible) is to promote social welfare. If so, we will favor welfare-promoting nudges. Consider, for example, a disclosure requirement for credit card companies, designed to promote informed choices, and likely to achieve that goal. Or suppose that we believe in individual autonomy and dignity. If so, we will favor nudges and choice architecture that promote those values. Consider, for example, an effort to prompt people to make their own choices about what kind of retirement plan they want, by asking them precisely that question when they begin employment.

If we value democratic self-government, we will be inclined to support nudges and choice architecture that can claim a democratic pedigree and that promote democratic goals. Any democracy has a form of choice architecture that helps define and constitute its own aspirations to self-government. Indeed, a Constitution can be seen as a kind of choice architecture for choice architects. A self-governing society might well nudge its citizens to participate in the political process and to vote. Political parties regularly engage in such nudging, and it is hardly illegitimate for public officials to encourage people to vote. In 2015, for example, Oregon adopted a system of automatic voter registration, a form of choice architecture that is unambiguously designed to promote participation in the political process through exercise of the franchise.⁶

Of course no one should approve of nudges or choice architecture in the abstract or as such. It remains possible that distrust of government, and faith in

Health Warnings on Packages (May 26, 2011), http://www.cdc.gov/media/releases/2011/p0526_cigarettewarnings.html. For a detailed account of the effects of such warnings, see Christine Jolls, *Debiasing Through Law and the First Amendment*, 67 STAN. L. REV. 1411 (2015).

5. Note as well that a disclosure requirement is a mandate, and no mere nudge, for the people or businesses on whom the requirement is imposed. It might be a nudge for consumers but a requirement for producers. I will say more about this point below.

6. See Cass R. Sunstein, *Oregon's Example For Voters Everywhere*, BLOOMBERG (Mar. 19, 2015), <http://www.bloombergview.com/articles/2015-03-19/oregon-s-example-for-voters-everywhere>.

markets, will lead us to minimize nudging on welfarist or autonomy grounds.⁷ Some nudges, and some forms of choice architecture, do indeed run into convincing ethical objections, above all because they undermine welfare, autonomy, or dignity. They might also compromise rather than promote self-government. Suppose, for example, that a nation establishes a default rule stating that unless voters explicitly indicate otherwise, they will be presumed to support the incumbent leader in the election. Or suppose that a nation establishes a default rule to the effect that unless citizens indicate otherwise, their estates will revert to the nation's most powerful political party upon their death. There is ample reason to question a default rule of this kind even if citizens are authorized to opt out. Some nudges have *illicit ends*, and they are objectionable for that reason.

There is also a pervasive question about manipulation. A nudge might preserve freedom of choice, but it might manipulate people and be objectionable for that reason. As we shall see, transparency and accountability are indispensable safeguards, and both nudges and choice architecture should be transparent.⁸ Even if so, there is a risk of manipulation, and that risk should be avoided. Many of the most interesting and complex ethical questions involve the disputed concept of manipulation, and I will devote some attention to that concept here.

I will be covering a great deal of territory, but three unifying themes should be kept in mind throughout. The first involves the relationship between nudges and human agency. The central point is that when nudges are in place, human agency is retained (because freedom of choice is not removed) and that agency always takes place in the context of some kind of choice architecture. The second general point involves the importance of having a sufficiently capacious sense of the category of nudges, and a full appreciation of the differences among them. Some nudges enlist or combat behavioral biases but others do not, and even among those that do enlist or combat such biases, there are significant differences. The third unifying theme is the need to bring ethical concerns in close contact with particular examples. A legitimate point about default rules may not apply to warnings or reminders. An ethical objection to the use of social norms may not apply to information disclosure. Here as elsewhere, abstraction can be a trap.

The remainder of this Article is organized as follows. Part II offers conceptual clarifications, designed to set the stage and to establish the inevitability of choice architecture. It contends that even spontaneous orders and invisible hands turn out to nudge, sometimes in extremely important ways. A

7. It is also possible, of course, that distrust of government and faith in markets will make us more inclined to minimize nudging on welfarist or autonomy grounds. See Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133 (2006).

8. Of course the term needs definition. At a minimum, I mean to suggest that no nudging should be hidden or free from public scrutiny. See *infra* Part III.C.

basic goal of the discussion is to establish the pointlessness of generalized objections to nudging, even from the state. Part III builds on this claim by exploring the diversity of nudges, the underlying psychology, and the relationship between nudges and navigability. It also emphasizes the need to distinguish between paternalistic nudges and nudges that respond to market failures, and also between educative (which are likely to promote autonomy) and noneducative nudges (which may or may not promote autonomy).

Part IV outlines the central criterion for paternalistic nudges, which is that they must improve people's welfare *as judged by themselves*. It also identifies conceptual and empirical difficulties with that criterion. Part V is the heart of the Article. Informed by the "as judged by themselves" criterion, it deals with the most important ethical objections to nudging, with particular emphasis on foundational ideas about welfare, autonomy, and dignity. Considerable discussion is also dedicated to the problem of manipulation.

II. Concepts and Definitions

A. In General

Nudges are interventions that steer people in particular directions but that also allow them to go their own way.⁹ A reminder is a nudge; so is a warning. A GPS nudges; a default rule nudges. To qualify as a nudge, an intervention must not impose significant material incentives (including disincentives).¹⁰ A subsidy is not a nudge; a tax is not a nudge; a fine or a jail sentence is not a nudge. To count as such, a nudge must fully preserve freedom of choice. If an intervention imposes significant material costs on choosers, it might of course be justified, but it is not a nudge.¹¹ Some nudges work because they inform people; other nudges work because they make certain choices easier; still other nudges work because of the power of inertia and procrastination.

When people make decisions, they do so against a background consisting of choice architecture.¹² A cafeteria has a design, and the design will affect what people choose. The same is true of websites. Department stores have architectures, and they can be designed so as to promote or discourage certain choices by shoppers (such as leaving without making a purchase). Even if the layout of a department store is a result of chance, or does not reflect the slightest effort to steer people, it will likely have consequences on what people end up

9. See RICHARD THALER & CASS R. SUNSTEIN, NUDGE 6 (2008) ("To count as a mere nudge, the intervention must be easy and *cheap to avoid*." (emphasis added). For numerous examples, see ANNUAL REPORT, *supra* note 9.

10. On some of the complexities here, see CASS R. SUNSTEIN, WHY NUDGE? 57-59 (2014).

11. See *id.*

12. See THALER & SUNSTEIN, *supra* note 9, at 3 ("A choice architect has the responsibility for organizing the context in which people make decisions.").

selecting.¹³ If, for example, people see certain items first, they are more likely to buy them.¹⁴

Both private and public institutions (including courts) create default rules. In fact they cannot dispense without them. The law of contract is permeated with default rules, which establish what happens if people do nothing. Consider, for example, implied warranties.¹⁵ If the contracting parties are silent on whether employment is “at will” or “for cause,” the law must supply a default; neither possibility is foreordained by nature or comes from the sky. People’s legal relationships with their employer, their mortgage provider, their rental car company, their credit card company, and even their spouse and their children consist in large part of default rules. A cell phone, a mortgage, a tablet, and a welfare program will inevitably come with defaults, which can be changed if the relevant people agree. Default rules nudge. They often operate like a GPS, or even help to shape preferences and values.

To some people, this is a disturbing or even threatening fact: They believe that people should be able to organize their lives as they like, and they are not at all enthusiastic about the idea that significant aspects of their lives are organized by default rules, which they did not themselves select, and which might well come from the practices, judgments, or wishes of other people. (As we shall see, many people object to nudges on the ground that they compromise individual agency.) Nonetheless, that organization is in place. Moreover, it is true that some default rules are a product of traditions, customs, spontaneous orders, and invisible hands – a comfort to some people who especially distrust public officials. But it would be extravagant to say that all (or most) of them are. And even if they are, they will nudge individuals who live with them, and it takes real work to transform them into law, where they will have significant effects.

Even if a default rule is chosen on the ground that it captures what most people will do, and is in that sense “market-mimicking,” it will likely have some effect on preferences and outcomes. A default rule establishes initial entitlements, and it can be important for that reason, influencing people’s preferences.¹⁶ For present purposes, the point is that default rules, of one or another kind, are sometimes unavoidable, or practically so.¹⁷

In addition, attention is a scarce resource. When applications (for loans, for educational opportunities, for refinancing mortgages, for training, for financial

13. See BRIAN WANSINK, *SLIM BY DESIGN* 116 (2014).

14. Eran Dayan & Maya Bar-Hillel, *Nudge to Nobesity II: Menu Positions Influence Food Orders*, 6 JUDGMENT AND DECISION MAKING 333 (2011).

15. See Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 MICH L. REV. 489, 516 (1989).

16. See Russell Korobkin, *The Endowment Effect and Legal Analysis*, 97 NW. L. REV. 1227, 1242 (2003) (“Evidence that the endowment effect exists is alone of great importance to legal scholars because it suggests that endowments will affect preferences and behaviors of individuals subject to the law.”).

17. An alternative, of course, is to call for active choosing, but sometimes that is not feasible. See generally Cass R. Sunstein, *Choosing Not to Choose*, 64 DUKE L.J. 1 (2014).

benefits of any kind) are complex and difficult, people may not apply; a great deal of money might be lost as a result.¹⁸ This point has implications for regulatory design. It suggests that the private sector may help or hurt people by focusing their attention in certain ways. The same is true for the public sector, whether or not it seeks to do so. A regulation might be written or applied in a way that makes certain features of a situation especially salient.

Spontaneous orders are celebrated by many people,¹⁹ and very reasonably so. If an order is spontaneous, in the sense that it reflects the voluntary, bottom-up decisions of those who created it, there is some reason to think that it reflects the judgments of many people about how it makes best sense to proceed. On certain assumptions, spontaneous orders can promote people's welfare,²⁰ but they are a form of choice architecture no less than intentional designs, and they will include a measure of nudging, not least if they create and perpetuate social norms.²¹ Invisible hands can nudge every bit as much as the most visible ones. To be sure, spontaneous order and invisible hands may be less dangerous than intentional designs, and on certain (controversial) assumptions²² they are likely to be benign (or better); but they are nonetheless forms of choice architecture.

For the future, we could imagine new forms of choice architecture that are designed to improve antipoverty programs²³; environmental programs²⁴; energy programs²⁵; retirement and social security programs²⁶; anti-obesity programs²⁷; educational programs²⁸; health care programs; and programs to increase organ donation.²⁹ We could also imagine forms of choice architecture that are designed to combat race and sex discrimination,³⁰ to help disabled people, and to promote

18. See Benjamin Keys et al., *Failure to Refinance* 1, 5 (Nat'l Bureau of Econ. Research, Working Paper No. 20798, 2014), <http://www.nber.org/papers/w20401>.

19. See FRIEDRICH HAYEK, FREEDOM, REASON, AND TRADITION 229 (1958).

20. See *id.*; see also ELINOR OSTROM, GOVERNING THE COMMONS 25-28 (1990).

21. See EDNA ULLMANN-MARGALIT, THE EMERGENCE OF NORMS 89 (1976).

22. Edna Ullmann-Margalit, *The Invisible Hand and the Cunning of Reason*, 64 SOC. RES. 181 (1997) (showing the assumptions on which invisible hands might be reliable and questioning those assumptions). For an extended exploration of why invisible hands might produce serious harm, see GEORGE AKERLOF & ROBERT SHILLER, PHISHING FOR PHOOLS (2015).

23. See SENDHIL MULLAINATHAN & ELDAR SHAFIR, SCARCITY 168 (2013).

24. Cass R. Sunstein & Lucia Reisch, *Automatically Green*, 38 HARV. ENV. L. REV. 128 (2014).

25. *Id.*

26. Ryan Bubb & Richard Pildes, *Why Behavioral Economics Trims its Sails*, HARV. L. REV. 1593, 1614 (2014).

27. See WANSINK, *supra* note 13, at 106.

28. See Adam Lavecchia et al., *Behavioral Economics of Education* 30 (Nat'l Bureau of Econ. Research, Working Paper No. 20609, 2014), <http://www.nber.org/papers/w20609>.

29. For an interesting empirical result, see Judd Kessler & Alvin Roth, *Don't Take 'No' for An Answer: An Experiment with Actual Organ Donor Registrations* 27 (Nat'l Bureau of Econ. Research, Working Paper No. 20378, 2014), <http://www.nber.org/papers/w20378> (finding that required active choosing is less effective than prompted choice at getting people to sign up for organ donation).

30. See Iris Bohnet et al., *When Performance Trumps Gender Bias: Joint Versus Separate Evaluation* 16 (Harvard Bus. Sch., Working Paper No. 12-083, 2012), <http://www.hbs.edu/faculty/Publication%20Files/12-083.pdf>.

economic growth. A great deal of future work needs to be devoted to choice architecture in these and related domains.³¹

There is no question that certain nudges, and certain kinds of choice architecture, can raise serious ethical problems.³² Consider, for example, a government that used nudges to promote discrimination on the basis of race, sex, or religion. Any fascist government might well (and almost certainly does) nudge. Terrorists nudge. Even truthful information (for example, about crime rates) might fan the flames of violence and prejudice. (If people learn that bad behavior is widespread, they might be more likely to engage in that behavior, because it is the social norm.)³³ Groups or nations that are committed to violence sometimes enlist nudges in their cause, often in the form of propaganda, which can be counted as a (bad) kind of nudge.³⁴ And even if nudges do not have illicit ends, it is possible to wonder whether those who enlist them are treating people with respect.

The most prominent concerns about nudging and choice architecture point to three foundational commitments: welfare, autonomy, and dignity. Some nudges could run afoul of one or more of these commitments. It is easy to identify welfare-reducing nudges that lead people to waste time or money³⁵; an unhelpful default rule for health insurance could fall into that category, as could an educational campaign designed to persuade people to purchase excessive insurance or to make foolish investments. Nudges could be, and often are, harmful to the environment.³⁶ Excessive pollution is, in part, a product of unhelpful choice architecture, because a wide range of default rules and legal permissions help make it possible.³⁷

B. The Inevitability of Choice Architecture

Consider in this light a tale from the novelist David Foster Wallace: “There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says ‘Morning, boys. How’s the

31. See WORLD BANK, WORLD DEVELOPMENT REPORT, MIND AND SOCIETY: HOW A BETTER UNDERSTANDING OF HUMAN BEHAVIOR CAN IMPROVE DEVELOPMENT POLICY 86 (2015). (exploring how behaviorally informed policies might promote development, in part by combating poverty).

32. See MARK WHITE, THE MANIPULATION OF CHOICE: ETHICS AND LIBERTARIAN PATERNALISM 81-102 (2013); REBONATO, *supra* note 1, at 153-87.

33. See THALER & SUNSTEIN, *supra* note 9, at 53-57.

34. See generally FRANK WESTERMAN, ENGINEERS OF THE SOUL: THE GRANDIOSE PROPAGANDA OF STALIN’S RUSSIA (2011).

35. See the weak effects of the opt-out design in Robert Letzler, *Knowing When To Quit: Default Choices, Demographics, and Fraud* 3 (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512276.

36. For relevant discussion, see Sunstein & Reisch, *supra* note 24, at 132.

37. See *id.*

water?’ And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes ‘What the hell is water?’”³⁸

This is a tale about choice architecture. Such architecture is inevitable, whether or not we see it. It is the equivalent of water. Weather is itself a form of choice architecture, because it influences what people decide.³⁹ Needless to say, human life is not imaginable without some kind of weather. Nature nudges. The common law is a regulatory system, and it will nudge, even if it allows people to have a great deal of flexibility.

In this light, choice architecture is inevitable. Human beings (or dogs or cats or horses) cannot wish it away. Any store has a design; some products are seen first, and others are not. Any menu places options at various locations. Television stations come with different numbers, and strikingly, numbers matter, even when the costs of switching are vanishingly low; people tend to choose the station at the lower number, so that channel 3 will obtain more viewers than channel 53.⁴⁰ A website has a design, which will affect what and whether people will choose.⁴¹

Nor can the state avoid nudging. Suppose that a government is or purports to be firmly committed to free markets, private property, and laissez-faire. Even so, it cannot simply refrain from acting—or from nudging. It creates its own choice architecture. As Hayek wrote, the task of establishing a competitive system provides “indeed a wide and unquestioned field for state activity,” for “in no system that could be rationally defended would the state just do nothing. An effective competitive system needs an intelligently designed and continuously adjusted legal framework as much as any other.”⁴²

As Hayek understood, a state that protects private property has to establish a set of prohibitions and permissions, including a set of default entitlements, establishing who has what before bargaining begins. Recall that the rules of contract (as well as property and tort) provide a form of choice architecture for social ordering. I have noted that initial entitlements and default rules nudge, because they affect people’s preferences, even if transaction costs are low and people can bargain as they see fit.⁴³ It remains true that choice architecture can maintain freedom of choice; it is also true that choice architects can at least aspire

38. David Foster Wallace, *In His Own Words*, INTELLIGENT LIFE (Sept. 19, 2008), <http://moreintelligentlife.com/story/david-foster-wallace-in-his-own-words>.

39. See Meghan R. Busse et al., *Projection Bias in the Car and Housing Markets 2* (Nat’l Bureau of Econ. Research, Working Paper No. 18212, 2014), <http://www.nber.org/papers/w18212>.

40. See Gregory Martin & Ali Yurukoglu, *Bias in Cable News: Real Effects and Polarization* (Nat’l Bureau of Econ. Research, Working Paper No. 20798, 2014), <http://www.nber.org/papers/w20798>.

41. STEVE KRUG, DON’T MAKE ME THINK REVISITED: A COMMON SENSE APPROACH TO WEB AND MOBILE USABILITY 10-19 (2014).

42. F.A. HAYEK, THE ROAD TO SERFDOM 40 (1943).

43. See Kevin M. Marzilli Ericson & Andreas Fuster, *The Endowment Effect 7* (Nat’l Bureau of Econ. Research, Working Paper No. 19384, 2013), <http://www.nber.org/papers/w19384>

to neutrality along important dimensions. But choice architecture itself is inevitable, which means that it is pointless to object to it on ethical grounds.

C. Spontaneous Orders and Visible Hands

At the same time, we can imagine the following view: Some forms of choice architecture and nudging are unavoidable, to be sure, but it is important and good if they are the product of nature or some kind of spontaneous order, rather than of conscious design, or of the action of any designer. Perhaps the law can build on that order; perhaps that law of contract, property, and tort do exactly that. Invisible-hand mechanisms⁴⁴ often produce choice architecture. On a time-honored view, much of law is in fact “customary law.” It codifies people’s actual practices, and it does not reflect any kind of dictation by public authorities. Consider Hayek’s celebration of the “empiricist, evolutionary tradition,” for which:

[T]he value of freedom consists mainly in the opportunity it provides for the growth of the undesigned, and the beneficial functioning of a free society rests largely on the existence of such freely grown institutions. There probably never has existed a genuine belief in freedom, and there certainly been no successful attempt to operate a free society, without a genuine reverence for grown institutions, for customs and habits⁴⁵

On this view, there is special reason, from the standpoint of freedom, for valuing forms of choice architecture that reflect the work of “grown institutions,” rather than designed ones. We might be comfortable with any nudging that reflects “customs and habits,” but suspicious of any nudging that displays no reverence for them. Here, then, is a foundation for skepticism about any kind of social engineering; the skepticism might be applied to nudges as well as to mandates and bans. We might be particularly concerned about governmentally-designed choice architecture, on the ground that public officials lack the information or the incentives to be trusted.

Even if the law of contract, property, and tort constitute forms of choice architecture, and even if it is not quite customary (and involves a degree of dictation and design), the relevant architecture can be made as flexible as possible and maintain a great deal of room for private ordering—and thus for freedom. To summarize a lengthy argument⁴⁶: the state, and the law, can provide the background rules for private interaction and decline to specify outcomes. Even if those rules turn out to nudge (as in the case of default rules), they are

44. For a superb discussion of invisible-hand explanations and their limitations, see Ullmann-Margalit, *supra* note 22.

45. F.A. HAYEK, THE CONSTITUTION OF LIBERTY: THE DEFINITIVE EDITION 122 (2013).

46. See generally FRIEDRICH HAYEK, THE CONSTITUTION OF LIBERTY (1976).

very different from social planning in all its forms—on one view, far more modest and less dangerous.

To be sure, the criminal law will include some dictation; no nation can deal with murder, assault, and rape with mere nudges. But perhaps the criminal law can restrict itself to prohibitions on force and fraud (and more generally play a role in correcting the standard market failures). A nation can certainly minimize the number of activities that it criminalizes. It might adopt a narrow account of the scope of the criminal law, perhaps focused on harm to others. And it might restrict any nudging to initiatives that operate in the service of the criminal law, narrowly conceived. For example, it might adopt public educational campaigns to discourage sexual violence.

On certain assumptions, much more in the way of self-conscious choice architecture by the state is especially worrisome, because it threatens welfare, autonomy, and dignity. But it is necessary to ask: What are those assumptions, and are they likely to be correct? Why and exactly when would spontaneous order be benign? (Is there some kind of social Darwinism behind enthusiasm for spontaneous order, reflecting a belief that voluntary choices and competition necessarily produce good results? That view is not easy to defend.)⁴⁷ It is true that action by government poses dangers and risks, but efforts to defend spontaneous orders and invisible hands run into well-understood problems and objections because there is no reason for confidence that the outcomes will be benign.⁴⁸

For example, a government that forbids discrimination on the basis of race and sex, or that takes steps to ensure reasonable accommodation of disabled people, is hardly relying on an invisible hand, and it is not drawing on custom. In any case, the argument for spontaneous orders seeks to restrict, above all, the coercive power of the state, not nudges as such. Whatever our theory of the legitimate domain of government, the most serious harms tend to come from mandates and bans (from genuine coercion), and not from nudges, which maintain freedom of choice.

It is true that spontaneous orders, invisible hands, and randomness can avoid some of the serious dangers, and some of the distinctive biases, that come from self-conscious nudging on the part of government.⁴⁹ If we are especially fearful of official mistakes—coming from incompetence or bad motivations—we will want to minimize the occasions for nudging.⁵⁰ And if we believe that invisible hand mechanisms generally promote welfare or freedom, we will not want to disturb their products, even if those products include nudges.

In my view, however, the strong position in favor of spontaneous orders and invisible hands cannot be defended. But my goal here is not to justify that

47. See AKERLOF & SHILLER, *supra* note 22, at 1-8.

48. *Id.*; Ullmann-Margalit, *supra* note 22, at 184-89 (1997).

49. See Edward Glaeser, *supra* note 7, at 136-39.

50. *Id.*

conclusion. The minimal point is that a degree of nudging by state officials cannot be avoided. If we are committed to spontaneous orders and invisible hands, we will be committed to a specific role for government, one that will include a specified choice architecture and specified nudges.

III. The Trap of Abstraction

To come to terms with the ethical questions, it is exceedingly important to bring first principles in contact with concrete practices. For purposes of orientation, it will be useful to give a more detailed accounting of potential nudges that might alter choice architecture.⁵¹ One reason is to avoid the trap of abstraction, which can create serious confusion when we are thinking about regulatory (or other) policy. As we shall see, the ethical evaluation of nudges depends on their concrete content, not on their status as nudges.

A. Nudges and Navigability

The most obvious nudges consist of default rules, which establish what happens if people do nothing at all.⁵² Others include simplification (for example, of applications for job training or financial aid); disclosure of factual information (for example, calorie labels); warnings, graphic or otherwise (for example, on cigarette packages); reminders (for example, of bills that are about to become due); increases in ease and convenience (for example, through website design); uses of social norms (for example, disclosure of how one's energy use compares to that of one's neighbors); nonmonetary rewards, such as public recognition; active choosing (as in the question: *what retirement plan do you want? or do you want to become an organ donor?*); and precommitment strategies⁵³ (through which people agree, in advance, to a particular course of conduct, such as a smoking cessation program).

It is important to acknowledge that some nudges preserve freedom of choice for a relevant population, while mandating action from some other population. Suppose, for example, that the government requires large employers to adopt automatic enrollment plans for either retirement or for health insurance.⁵⁴ If so, employees are nudged, but employers are coerced. Or suppose that the government requires chain restaurants (including movie theaters) to display

51. See Cass R. Sunstein, *Nudging: A Very Short Guide*, 3 J. CONSUMER POL'Y 583 (2014). For a set of nudges—none of which is easily subject to ethical objections—see ANNUAL REPORT, *supra* note 9.

52. See Eric Johnson et al., *Defaults, Framing and Privacy: Why Opting In-Opting Out*, 13 MARKETING LETTERS 5 (2002).

53. See generally IAN AYRES, *CARROTS AND STICKS: UNLOCK THE POWERS OF INCENTIVES TO GET THINGS DONE* (2011) (reviewing the power of self-commitment devices to change people's behaviors).

54. The Affordable Care Act, in fact, does this.

calories to consumers.⁵⁵ If so, customers are nudged, but restaurants are coerced. Some nudges from government take the form of requiring some group *X* to nudge some other group *Y*. This point should be kept in mind in evaluating ethical objections, because coercion raises distinctive concerns, and sometimes government coerces the private sector to nudge consumers.

In behavioral science, it has become standard to distinguish between two families of cognitive operations: System 1, which is fast, automatic, and intuitive, and System 2, which is slow, calculative, and deliberative.⁵⁶ System 2 can and does err, but System 1 is distinctly associated with identifiable behavioral biases.⁵⁷ To be sure, there is, in some circles, intense controversy about the appropriate evaluation of the automatic system and about the extent to which it should be associated with error. Some people argue that our intuitions usually work well in the situations in which we ordinarily find ourselves.⁵⁸ But there is no question that our intuitions often misfire, and that a good nudge could provide a great deal of help.

Some nudges, imposed by regulatory agencies, attempt to counteract System 1 and to strengthen the hand of System 2 by improving the role of deliberation and people's considered judgments—as, for example, through disclosure of relevant information, debiasing,⁵⁹ and the use of precommitment strategies. Other nudges are designed to appeal to, or to activate, System 1—as in the cases of graphic warnings.⁶⁰ Some nudges do not appeal to System 1, but work because of its operation—as, for example, where default rules have large effects because of the power of inertia.⁶¹

A nudge might be justified on the ground that it helps counteract a behavioral bias,⁶² and (as we shall see) some people object to such efforts, especially if they seem to target or to exploit System 1. But (and this is an

55. The Affordable Care Act does this as well, for which FDA rules were recently finalized. See Press Release, FDA, FDA Finalizes Menu and Vending Machine Calorie Labelling Rules (Nov. 25, 2014), <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm423952.htm>.

56. See DANIEL KAHNEMAN, THINKING, FAST AND SLOW 20 (2011).

57. See *id.* at 415-16.

58. This position is vigorously defended throughout GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999); for a general discussion, see MARK KELMAN, THE HEURISTICS DEBATE 19 (2011). In my view, the outcome of this debate of this occasionally heated (and somewhat theological) debate usually does not have strong implications for policy, practice, or ethics. Everyone should agree that heuristics generally work well; that is why they exist. Everyone should also agree that in important cases, boundedly rational people make mistakes. When they make mistakes, some kind of nudge might help. To be sure, the best nudge may or may not involve education. See *infra* p. 124.

59. See generally Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, 35 J. LEGAL STUD. 199 (2006).

60. For an example of a recent treatment of mandatory graphic warnings, see R.J. Reynolds Tobacco Co. v. U.S. Food & Drug Admin., 823 F. Supp. 2d 36 (D.D.C. 2011), *aff'd*, 696 F.3d 1205 (D.C. Cir. 2012).

61. See Johnson et al., *supra* note 52.

62. See Jolls & Sunstein, *supra* note 59, at 227.

important point) a behavioral bias is *not* a necessary justification for a nudge,⁶³ and nudges need not target or exploit System 1 in any way. Disclosure of information can be helpful even in the absence of any bias. A default rule simplifies life and might therefore be desirable whether or not a behavioral bias is involved. A GPS is useful even for people who do not suffer from a cognitive bias affecting their ability to drive to their desired destination.⁶⁴

As the GPS example suggests, many nudges have the goal of *increasing navigability*—of making it easier for people to get to their preferred destination. Such nudges stem from an understanding that life can be simple or hard to navigate, and a goal of helpful choice architecture is to promote simpler navigation. To date, there has been far too little attention to the close relationship between navigability and (good) nudges. Insofar as the goal is to promote navigability, the ethical objections are greatly weakened and might well dissipate.

Nudges can have a substantial effect on both individual lives and social welfare. In Denmark, for example, automatic enrollment in retirement plans has had a much larger effect than substantial tax incentives.⁶⁵ In the United States, efforts to inform consumers of how their energy use compares to that of their neighbors has had the same (significant) effect on household energy use as a significant spike in the short-term cost of electricity.⁶⁶ Simplification of the financial aid form, to assist people who seek to attend college, has been found to have as large an effect in promoting college attendance as a several thousand dollar increase in financial aid.⁶⁷

B. Three Distinctions

For purposes of evaluating the ethical questions, three distinctions are particularly important, because they help clarify the underlying concerns. First, *paternalistic nudges* should be distinguished from *market failure nudges*. Some of the most familiar nudges are designed to protect people from their own mistakes (including a behavioral bias); consider a default rule designed to

63. This conclusion means that the highly illuminating discussion in REBONATO, *supra* note 1, offers a mistaken definition.

64. Sarah Conly writes that for those who endorse nudging, “[t]he assumption is that because our decision-making ability is limited we need to use nonrational means to seduce people into doing what is good for them, and are trying to get people to act through the use of nonrational means.” SARAH CONLY, *AGAINST AUTONOMY* 30 (2012). This is not the assumption that lies behind nudging, though perhaps some nudges can be understood in this way.

65. Raj Chetty et al., *Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark* 38 (Nat’l Bureau of Econ. Research, Working Paper No. 18565, 2012), <http://www.nber.org/papers/w18565>

66. See Hunt Allcott, *Social Norms and Energy Conservation*, 95 J. PUB. ECON. 1082, 1082 (2011).

67. See Eric Bettinger et al., *The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experience* (Nat’l Bureau of Econ. Research Working Paper No. 15361, 2009), <http://www.nber.org/papers/w15361>.

increase savings. Other nudges are designed to respond to a standard kind of market failure, which can arise even if individuals are making no mistake at all (and so paternalism is not an issue). Consider a default rule designed to reduce pollution (as, for example, in the form of a default rule in favor of green energy).

In the category of market failure nudges, we can identify *externality-reducing nudges*, *prisoner's dilemma nudges*, and *coordination*⁶⁸ nudges. For nudges that fall into these categories, the governing question should be: Do they increase social welfare, rightly understood?⁶⁹ Cost-benefit analysis is the best available way of operationalizing that question, though it has significant gaps and limitations, not least if distributional considerations turn out to be relevant.⁷⁰

There is broad agreement that it is legitimate for government to respond to market failures.⁷¹ If the government is trying to reduce a collective action problem that produces high levels of pollution, it does not raise the kinds of ethical concerns that come into play if the government is acting paternalistically. It follows that market failure nudges should not be especially controversial in principle, though we might well worry over questions of effectiveness.⁷² In the face of a standard market failure, a mere nudge is usually not enough; coercion might well be justified (perhaps in the form of a corrective tax, perhaps in the form of a regulatory mandate). But a nudge might prove to be complementary to coercion, and in some ways, it might be a substitute.⁷³

Second, *educative nudges* should be distinguished from nudges that *lack educative features*. Educative nudges attempt to inform people, so that they can make better choices for themselves. Other nudges are meant to help people without increasing their knowledge or understanding; default rules have this characteristic. If the focus is on *increasing people's own powers of agency*, educative nudges should not be especially controversial on ethical grounds, and indeed they should be welcomed—though their benefits might not justify their costs,⁷⁴ and though they can also run into problems of effectiveness.⁷⁵

Third, *nudges that enlist or exploit behavioral biases* should be distinguished from nudges that do no such thing. In particular, we have seen that some nudges enlist or exploit System 1 whereas other nudges appeal to System

68. On coordination, with implications for productive nudges, see Edna Ullmann-Margalit, *Coordination Norms and Social Choice*, 11 ERKENNTNIS 143 (1977).

69. Of course there is a great deal of dispute about how social welfare is rightly understood. For a valuable discussion, see MATTHEW D. ADLER, *WELL-BEING AND FAIR DISTRIBUTION* (2012).

70. See Sunstein & Reisch, *supra* note 24, at 149.

71. See STEPHEN BREYER, *REGULATION AND ITS REFORM* 13-35 (1984)..

72. See Bubb & Pildes, *supra* note 26, at 1602; Sunstein & Reisch, *supra* note 24, at

148

73. See Sunstein & Reisch, *supra* note 24 at 156.

74. See George Loewenstein et al., *Disclosure: Psychology Changes Everything*, 6

ANN. REV. OF ECON. 391, 392 (2014).

75. See Lauren Willis, *The Financial Education Fallacy*, 101 AM. ECON. REV. 429,

430 (2011).

2.⁷⁶ Efforts to target, or to benefit from, behavioral biases tend to be more controversial, on ethical grounds, than efforts to appeal to deliberative capacities.⁷⁷ The reason is that the former appear to be more manipulative and less respectful of people's capacity for agency (a point to which I will return).

It follows that the most controversial nudges are paternalistic, noneducative, and designed to enlist or exploit behavioral biases.⁷⁸

C. Illicit Reasons and Transparency

It must be acknowledged that choice architecture can be altered or that new nudges can be introduced for illicit reasons. Indeed many of the most powerful objections to nudges, and to changes in choice architecture, are based on a judgment that the underlying motivations are illicit.⁷⁹ With these points, there is no objection to nudges as such; the objection is to the grounds for the particular nudges.

For example, an imaginable default rule might skew the democratic process by saying that voters are presumed to vote for the incumbent politician, unless they specify otherwise. Such a rule would violate principles of neutrality that are implicit in democratic norms; it would be unacceptable for that reason. Alternatively, a warning might try to frighten people by fabricating stories about the supposedly nefarious plans of members of a minority group. Social norms might be invoked to encourage people to buy unhealthy products ("most people are buying these products; you should too!"). In extreme cases, private or public institutions might try to nudge people toward violence.

It must also be acknowledged that the best choice architecture often calls for active choosing.⁸⁰ Sometimes the right approach is to *require* people to choose, so as to ensure that their will is actually expressed. Sometimes it is best to *prompt* choice, by asking people what they want, without imposing any requirement that they do so.⁸¹ A prompt is emphatically a nudge, designed to get people to express their will, and it might be unaccompanied by any effort to steer people in a preferred direction—except in the direction of choosing.

Choice architecture should be transparent and subject to public scrutiny, certainly if public officials are responsible for it. At a minimum, this proposition means that when such officials institute some kind of reform, they must not hide it from the public. If officials issue a regulation changing a default rule so as to promote clean energy or conservation, they should disclose what they are doing,

76. See KAHNEMAN, *supra* note 56, at 413-14.

77. REBONATO, *supra* note 1, at 91.

78. For a detailed discussion of public reactions to nudges, see Cass R. Sunstein, *Do People Like Nudges?* (2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2604084 (assessing people's preferences toward nudges with a nationally representative survey).

79. See Glaeser, *supra* note 7, at 148-55; REBONATO, *supra* note 1, at 103.

80. See Cass R. Sunstein, *Deciding By Default*, 162 U. PA. L. REV. 1, 42-46 (2013).

81. Kessler & Roth, *supra* note 29, at 28.

and they should also explain why they are doing it (with reference to any behavioral bias, if that is the reason for the change). Self-government itself requires public scrutiny of nudges. Such scrutiny is an important *ex ante* safeguard against harmful nudges; it is also an important *ex post* corrective. Transparency and public scrutiny can reduce the likelihood of welfare-reducing choice architecture. Nations should also treat their citizens with respect, and allowing public scrutiny shows a measure of respect and simultaneously reduces the risk that nudges will reduce welfare, or intrude on autonomy or dignity.

There is a question whether transparency and public scrutiny are sufficient rather than merely necessary. The answer is that they are not sufficient. We could imagine forms of choice architecture that would be unacceptable even if they were fully transparent; consider (transparent) architecture designed to entrench inequality on the basis of sex. Here again, the problem is that the goals of the relevant nudge are illicit. As we shall see, it is also possible to imagine cases of manipulation, in which the goals are not illicit, but in which the fact of transparency might not be sufficient to justify a nudge. A transparent nudge, announced in advance but taking the form of subliminal advertising of one or another kind, would run into legitimate objections about manipulation.

IV. "As Judged By Themselves"

A. *The Basic Standard*

Let us now turn to nudges that are not designed to address a standard market failure and that are intended to help people to avoid their own mistakes. When third parties are not at risk, and when the welfare of choosers is all that is involved, the central objective of nudging is to "influence choices in a way that will make choosers better off, *as judged by themselves*."⁸² In many cases, that standard is straightforward to apply. If a GPS steers people toward a destination that is not their own, it is not working well. And if it offers them a longer and less convenient route, it will not make choosers better off by their own lights (unless that is the route they want, perhaps because it is scenic).

Many nudges can be evaluated under the "as judged by themselves" standard; consider a reminder, a warning, a default rule, or disclosure of relevant information. To see whether the standard is met, we would have to take each nudge on its own. But the standard will often provide sufficient guidance.

This understanding of the objective of nudging should go some way toward defusing the principal ethical concerns about nudging. If the choice architect is actually succeeding in making choosers better off by their own lights, there would seem to be no objection from the standpoint of welfare. Nor is it clear that there is a problem from the standpoint of autonomy or dignity. I will return to these points in Part V.

82. THALER & SUNSTEIN, *supra* note 9, at 5.

B. Questions and Doubts

At the same time, it must be acknowledged that the standard raises normative, conceptual, and empirical challenges.

1. Objectively Good Lives?

Some people believe that human lives can be objectively good or objectively bad, and that choosers can and do make objective mistakes about what makes their lives good. As they are understood in political philosophy, “perfectionist”⁸³ approaches emphasize the importance and legitimacy of increasing the likelihood that people will have lives that are actually good, rather than of respecting multiple conceptions of the good. (Of course there are different forms of perfectionism, and I am bracketing many complexities here.)

For people who have this belief, the “as judged by themselves” standard is based on a fundamental mistake, which is that it allows the subjective judgments of choosers to prevail even if they are objectively wrong. Imagine, for example, that a chooser makes decisions that ensure a life that is short and unhealthy, or that is without either meaning or pleasure,⁸⁴ or that involves a great deal of suffering. It might be asked: Why should choice architects defer to choosers in such circumstances?

This question raises serious questions within political philosophy, which I cannot answer here.⁸⁵ To the extent that choice architects defer to choosers, it might be because of a moral judgment that choosers have ultimate sovereignty over their own lives, or it might be because of their own humility—their understanding that in general, and notwithstanding the existence of behavioral biases, they might well have epistemic disadvantages as compared with those whose own lives are at stake. For present purposes, the central point is that insofar as choice architects adopt the “as judged by themselves” standard, they reject perfectionism, and they do so not inadvertently but on principle.

2. Ex Ante or Ex Post?

The “as judged by themselves” standard raises this question: Do we ask about choosers’ judgments before the nudge, or instead after? Choosers’ ex ante judgments might diverge from their ex post judgments. If choosers’ judgments are constructed by the nudge, then *choice architects might be engineering the*

83. A form of liberal perfectionism is defended in JOSEPH RAZ, *THE MORALITY OF FREEDOM* (1985). Perfectionism is criticized in CONLY, *supra* note 64, at 100-25.

84. On the importance of purpose and pleasure, see PAUL DOLAN, *HAPPINESS BY DESIGN* 34 (2014).

85. Relevant discussion can be found in ADLER, *supra* note 69; MARTHA NUSSBAUM, *CREATING CAPABILITIES* (2013); AMARTYA SEN, *COMMODITIES AND CAPABILITIES* (1999). For a short, vivid set of objections to perfectionism, see CONLY, *supra* note 64, at 109-25.

very judgment from which they are claiming authority. That is a serious problem for the “as judged by themselves” standard.

Suppose, for example, that with a “green” default rule—one that provides an environmentally friendly energy provider, subject to opt out in favor of a cheaper but environmentally inferior provider—choosers (or a majority) are perfectly content. But suppose that they (or a majority) would also be content with the opposite default rule. Which judgments matter? Wherever the nudge influences choosers’ judgments, that question raises serious puzzles.

In some cases, the nudge is unlikely to affect choosers’ judgments; they will be the same *ex ante* and *ex post*. Whatever the default rule, most people are not likely to want to devote 50 percent of their salary to savings, or to have a health insurance policy that does not pay for surgeries, or to have meals that consist of old salami slices on stale bread. But when *ex post* and *ex ante* judgments differ, the standard becomes more difficult to apply.⁸⁶ One option would be to use active choosing to see what people actually want. Another would be to explore the number of opt-outs under different default rules.⁸⁷ A third would be to attempt a more direct inquiry into people’s welfare under different forms of choice architecture, though admittedly any such inquiry raises challenges of its own.⁸⁸

3. Preferences About Preferences

An additional question is raised by the fact that people do not only have preferences (or first-order preferences); they also have preferences about their preferences (or second-order preferences).⁸⁹ People might want to eat delicious but fattening foods, or to spend most of their monthly salary every month, but they might not want to want those things. In applying the “as judged by themselves standard,” should choice architects consult first-order or second-order preferences?

Some imaginable cases are difficult, but in general, the answer is straightforward: If second-order preferences reflect System 2 thinking—understood, in this context, as people’s reflective judgments as opposed to their momentary impulses—there is a strong argument that such preferences have authority.⁹⁰ To be sure, System 1 has its claims. If people greatly enjoy certain activities, their enjoyment should not be disparaged; it is part of what makes life

86. See Jacob Goldin, *Which Way To Nudge? Uncovering Preferences in the Behavioral Age*, 125 YALE L.J. (forthcoming 2015).

87. THALER & SUNSTEIN, *supra* note 9.

88. See DOLAN, *supra* note 84. Goldin, *supra* note 86, offers a complex and ingenious approach to this problem, which is worth careful consultation.

89. See Harry Frankfurt, *Freedom of the Will and the Concept of a Person*, 68 J. PHIL. 5 (1971).

90. I am bracketing some complexities here. For example, people’s intuitions might be well-trained, and their reflective judgments might be unreliable. We could imagine a tennis player who knows exactly what to do when a first serve goes to his backhand, even though he would not do very well in giving a reflective explanation, to himself or to others, what to do in those circumstances.

worth living. If a nudge discourages people from eating foods that they enjoy, or from engaging in activities that they love, it is imposing a genuine welfare loss. But if people think, on reflection, that the nudge is directing them in a way that they endorse, then the “as judged by themselves” standard is met.⁹¹

4. Informed Judgments and Empirical Puzzles

When we ask about choosers’ judgments, what kind of information do we expect choosers to have? It makes sense to say that choice architects should defer to choosers’ informed judgments, rather than their uninformed ones. But if choice architects do not focus on choosers’ actual judgments, and instead ask what choosers would do if they were informed, there is a risk that choice architects will be relying on their own values and beliefs, rather than choosers’ own. In any case, such architects might lack sufficient information to know whether informed choosers deem themselves to be better off. It might not be at all simple for outsiders to compare (from the point of view of informed choosers) the various outcomes that stem from different nudges.

In some cases, these points might raise serious conceptual and empirical challenges. Nonetheless, the idea of choosers’ informed judgments serves as the lodestar, and it imposes real discipline.⁹² Certainly choice architects should be focused on the welfare of choosers, rather than their own. (In a well-functioning market system, that focus is essentially guaranteed for market participants, at least under optimistic assumptions.)⁹³

5. Self-Control

There are also hard questions about how to handle the “as judged by themselves” standard in the face of self-control problems. Suppose that someone faces such problems and is aware of that fact—but nonetheless wishes, at Time 1, to give into his impulses. Do we look to the assessment of (1) the alcoholic, who really wants that beer, (2) the would-be former alcoholic, who wants to quit, or (3) the actual former alcoholic, who is grateful to have been nudged away from alcoholism? In some ways, this question replicates those involving *ex ante*

91. There are, however, some difficult questions raised by time inconsistency. Suppose that at Time 1, Edward wants to engage in some behavior, and does, and enjoys a welfare gain, but at Time 2, Edward wishes that his former self had not engaged in that behavior, and experiences a welfare loss. If so, “as judged by themselves” standard runs into genuine challenges. See Douglas Whitman & Mario Rizzo, *The Problematic Welfare Standards of Behavioral Paternalism*, REV. PHIL. & PSYCHOL. (forthcoming 2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2560976

92. As noted, difficult questions might be raised where subjective well-being departs from objective well-being; I am bracketing those questions. See NUSSBAUM, *supra* note 85;; AMARTYA SEN, *DEVELOPMENT AS FREEDOM* (2009). Note also that in the presence of affective forecasting errors, an emphasis on subjective well-being leads to a focus on what, in fact, makes people (subjectively) better off, not on what they anticipate will make them better off.

93. Optimistic, not realistic. See OREN BAR-GILL, *SEDUCTION BY CONTRACT* 6-43 (2012); AKERLOF & SHILLER, *supra* note 22.

versus *ex post* judgments, and also those involving preferences about preferences. But insofar as the focus is on self-control problems, the issue is distinctive.

It is reasonable to insist that most former alcoholics are glad that they are no longer alcoholics (at least in the absence of highly unusual circumstances). For that reason, there is a strong argument that the “as judged by themselves” criterion should be taken to refer to the judgment of the person who is no longer in the grip of an addiction. That proposition is sufficient to resolve many real-world cases. Nonetheless, there can be a thin line between a self-control problem and a legitimate focus on short-term pleasure; that question deserves more extended treatment. No choice architect should engage in a program of nudging that disregards the importance of short-term pleasures, or pleasures in general, which are of course crucial parts of good lives.⁹⁴

V. Applied Ethics

I have suggested that the principal ethical objections to nudges point to welfare, autonomy, and dignity. I now turn to those objections. Although it raises all three concerns, the topic of manipulation will be given a separate treatment in this Part. I shall also offer a few words about the risk of ignorant or biased choice architects.

A. Welfare

Choice architecture may or may not be paternalistic. But it is true that nudges can be seen as a form of “libertarian paternalism” insofar as they attempt to use choice architecture to steer choosers in directions that will promote their welfare (again, as judged by choosers themselves).⁹⁵ If we believe that choosers know best, any such effort at steering might seem misguided, at least if welfare is our lodestar.

To evaluate this objection, it is important to see that we are speaking here of a distinctive form of paternalism in the sense that it is at once a) soft and b) means-oriented.⁹⁶ It is soft insofar as it avoids coercion or material incentives and thus fully maintains freedom of choice.⁹⁷ It is means-oriented insofar as it does not attempt to question or alter people’s ends. Like a GPS, it respects those ends (subject to the various complexities discussed above). To those who object to paternalism on welfare grounds, the most serious concerns arise in the face of coercion (where freedom of choice is blocked) and when social planners, or

94. See DOLAN, *supra* note 84, at 34.

95. THALER & SUNSTEIN, *supra* note 9, at 5.

96. See SUNSTEIN, *supra* note 10, at 20.

97. On some of the complexities here, see *id.*

choice architects, do not respect people's ends.⁹⁸ To this extent, nudges aspire to avoid some of the standard welfarist objections to paternalism.⁹⁹

Nonetheless, some skeptics, concerned about welfare, object to paternalism as such.¹⁰⁰ Perhaps people are the best judges not only of their ends, but also of the best means to achieve those ends, given their own tastes and values. (People might reject the route suggested by the GPS on the ground that they prefer the scenic alternative; the GPS might not easily capture or serve their ends.) Moreover, the distinction between means and ends is not always simple and straightforward. One question is the level of abstraction at which we describe people's ends. If we describe people's ends at a level of great specificity—eating that brownie, having that cigarette, texting while driving—then people's means effectively *are* their ends. The brownie is exactly what they want; it is not a means to anything at all (except the experience of eating it).

If, by contrast, we describe people's ends at a level of high abstraction—"having a good life"—then nearly everything is a means to those ends. But if we do that, then we will not be capturing people's actual concerns; we will be disregarding what matters to them. These points do raise some problems for those who favor a solely means-oriented form of paternalism. They must be careful to ensure that they are not describing people's ends at a sufficiently high level of abstraction as to misconceive what people care about.¹⁰¹

But insofar as a GPS is a guiding analogy, it is not easy to see nudges as objectionable on welfare grounds. Many nudges are entirely focused on helping people to identify the best means for achieving their own preferred ends. Consider cases in which people are *mistaken about facts* (with respect to the characteristics of, say, a consumer product or an investment). If a nudge informs them, then it is respecting their ends and is likely to promote their welfare. Or suppose that certain product characteristics are in some sense shrouded, and the nudge helps people to see them for what they are. Or suppose that people suffer from a behavioral bias—perhaps because they use the availability heuristic, perhaps because of unrealistic optimism. A nudge that corrects their mistake can help them to achieve their ends.

To be sure, some behavioral biases are not easy to analyze in these terms. If people suffer from present bias, is a nudge a form of paternalism about means? Suppose that people eat high calorie food, or drink a great deal, or fail to exercise, because they value today and tomorrow, and not so much next year or next

98. This is the fundamental concern in John Stuart Mill, *On Liberty*, in THE BASIC WRITINGS OF JOHN STUART MILL: ON LIBERTY, THE SUBJECTION OF WOMEN, AND UTILITARIANISM 3, 11-12 (2002).

99. The debate over perfectionism is of course relevant here. See CONLY, *supra* note 64.

100. See, e.g., REBONATO, *supra* note 1; Joshua D. Wright & Douglas H. Ginsburg, *Behavioral Law and Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 NW. U. L. REV. 1033 (2012).

101. See REBONATO, *supra* note 1, at 56-57.

decade. If a nudge succeeds in getting people to focus on their long-term interests, it might increase aggregate (intrapersonal) welfare over time. But is such a nudge focused solely on means? If a person is seen a series of selves extending over time, the choice architect is effectively redistributing welfare from earlier selves to later ones (and possibly maximizing people's lifetime welfare as well). But it is not clear that we can speak, in such cases, of means paternalism. And if a person is seen as continuous over time, and not a series of selves, efforts to counteract present bias are, by hypothesis, undermining the ends of the chooser at the time of choice.

A principal reason for regulators and other policymakers to reject any form of paternalism involves welfare: Perhaps people are the best judges of what will promote their interests, and perhaps outsiders will blunder (as John Stuart Mill believed).¹⁰² Consider Hayek's remarkable suggestion that "the awareness of our irremediable ignorance of most of what is known to somebody [who is a planner] is the chief basis of the argument for liberty."¹⁰³ A form of paternalism that maintains freedom of choice, and that is focused on means, is less likely to be objectionable on welfare grounds, certainly if we attend to behavioral biases.

In fact it is even possible that welfarists should ultimately embrace coercive paternalism, at least in the face of such biases.¹⁰⁴ When paternalism would improve welfare, welfarists should support paternalism. For welfarists, paternalism should be evaluated on a case-by-case basis – unless there is some systematic reason to support a presumption against paternalism.¹⁰⁵

Perhaps there is good reason for such a presumption, rooted in a judgment that choosers are likely to have better information than choice architects.¹⁰⁶ But in some cases, that judgment is incorrect, because choosers lack knowledge of facts. Information-providing nudges are a natural corrective. In some cases, a good default rule—say, automatic enrollment in pension programs—is hard to reject on welfarist grounds. To be sure, active choosing might be better than a default rule, but that conclusion is not obvious; active choosing might impose real burdens on choosers (who would prefer not to choose) and might not reduce errors.¹⁰⁷ Welfarists might well be inclined to favor choice-preserving approaches, on the theory that individuals usually well know what best fits their circumstances, but if welfare is our guide, the fact that a default rule has a paternalistic dimension should not be decisive against it.

102. JOHN STUART MILL, ON LIBERTY 8 (Kathy Casey ed., 2002) (1859).

103. FRIEDRICH HAYEK, THE MARKET AND OTHER ORDERS, in THE COLLECTED WORKS OF F.A. HAYEK 384 (Bruce Caldwell ed., 2013).

104. CONLY, *supra* note 64, at 32.

105. See Cass R. Sunstein, *Nudges v. Shoves*, 127 HARV. L. REV. FORUM 210 (2014).

106. See *id.* at 211. For another, nonwelfarist view of what is wrong with paternalism, see Nicholas Cornell, *A Third Theory of Paternalism*, 113 MICH. L. REV. 1295 (2015) (arguing that paternalism may be wrong because it implies, and expresses a judgment, that the chooser is incapable of making good choices for herself).

107. See CASS R. SUNSTEIN, CHOOSING NOT TO CHOOSE: UNDERSTANDING THE VALUE OF CHOICE 144-149 (2015).

But consider here a different kind of concern, one that also grows from a focus on welfare: *Choice-making is a muscle, and the ability to choose well can be strengthened through exercise.* If nudges would make the muscle atrophy, we would have an argument against them, because people's welfare depends on muscle-strengthening. We could imagine an ethical objection that would contend that some nudges do not allow people to build up their own capacities, and might even undermine their incentive to do so. If so, people's welfare is reduced (and it should be easy to see that there is an autonomy concern as well).

Here too, it is necessary to investigate the particulars—the kinds of nudges and choice architecture that are involved. Active choosing and prompted choice hardly impede learning. Nor do information and reminders. On the contrary, they promote learning. Nudges of this kind exercise the choice-making muscle, rather than the opposite.¹⁰⁸

With respect to learning, a potential problem does come from default rules. It is possible to say that active choosing is far better than defaults, simply because choosing may promote learning. Consider, for example, the question whether employers should ask employees to make active choices about their retirement plans, or whether they should instead default people into plans that fit their situations. The potential for learning might well count in favor of active choosing; if employees must choose, and if they are helped or encouraged to learn, the resulting knowledge might stand them in good stead for the rest of their lives.¹⁰⁹ If people are defaulted into certain outcomes, they do not add to their stock of knowledge, and that may be a significant lost opportunity.

But the argument for learning depends on the setting. (Recall the earlier discussion of educative nudges.) There are domains in which the choice-making muscle needs rest rather than exercise. For most people, it is not important to become experts in the numerous decisions that lead to default settings in cell phones, and hence the use of such settings is not objectionable. The same point holds in many other contexts in which institutions rely on defaults rather than active choosing. Retirement savings might itself be such a context. To know whether choice architects should opt for active choosing, it is necessary to explore whether the context is one in which it is valuable, all things considered, for choosers to acquire a stock of knowledge. It is not necessarily terrible if employers create default allocations for retirement plans, so long as those allocations are in the interest of all or most employees.

It is true, of course, that education well might be the appropriate response if people are likely to make mistakes (whether they lack information or suffer from some kind of behavioral bias). Much of the time, the first and the best line of defense is education, which might itself be characterized as a nudge, and

108. See Benjamin York & Susanna Loeb, *One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers* 31 (Nat'l Bureau of Econ. Research, Working Paper No. 20659, 2014), <http://www.nber.org/papers/w20659>.

109. See Sunstein, *supra* note 17.

which certainly counts as a form of choice architecture. We have seen that educative nudges are an important part of the repertoire of the choice architect. Invoking this point, Jeremy Waldron writes: "I wish, though, that I could be made a better chooser rather than having someone on high take advantage (even for my own benefit) of my current thoughtlessness and my shabby intuitions."¹¹⁰

But education has its limits. People benefit from default rules with respect to cell phones, tablets, health insurance policies, and rental car agreements. (To be sure, not all such rules are beneficial.) If people had to obtain sufficient education on all of the underlying issues, they would quickly run out of time. On welfare grounds, a default rule is often the best approach, because it would preserve desirable outcomes (again, from the standpoint of choosers themselves) without requiring people to take the functional equivalent of a course in, say, statistics or finance.¹¹¹

For the welfarist, there is a recurring question whether, in particular circumstances, the costs of education justify the benefits. For those who are engaged in many activities, it would be impossibly demanding to insist on the kind of education that would allow active choices about all relevant features. Default rules may well be best. Everything depends on the facts, but if welfare is our guide, there is a good argument that default rules are preferable to financial education with respect to important retirement issues.¹¹²

B. Autonomy

An independent reason to reject paternalism involves autonomy and the idea of respect for persons. Stephen Darwall writes that the

[O]bjectionable character of paternalism of this sort is not that those who seek to benefit us against our wishes are likely to be wrong about what really benefits us. . . . It is, rather, primarily a failure of respect, a failure to recognize the authority that persons have to demand, within certain limits, that they be allowed to make their own choices for themselves.¹¹³

Do nudges intrude on autonomy? The answer depends on what kind of nudge is involved. Many nudges promote autonomy. A choice should not be considered autonomous if it is based on ignorance; if someone takes a stomach medicine in the belief that it will help with a cold, there is no interference with autonomy in informing him that it is stomach medicine. Autonomy requires

110. See Jeremy Waldron, *It's All for Your Own Good*, NEW YORK REVIEW OF BOOKS (2014), <http://www.nybooks.com/articles/archives/2014/oct/09/cass-sunstein-its-all-your-own-good/>.

111. See Willis, *supra* note 75, at 429.

112. See *id.* at 432.

113. See Stephen Darwall, *The Value of Autonomy and the Autonomy of the Will*, 116 ETHICS 263, 268 (2006).

informed choices, and many nudges are designed specifically to ensure that choices are informed.¹¹⁴

In the face of a behavioral bias, or some kind of systematic mistake (by the actor's own lights), it is hardly clear that a nudge infringes on autonomy, rightly understood.¹¹⁵ When nudges help correct some kind of bias, they might well promote people's autonomy—most obviously when they help people to have a better understanding of the facts. Suppose that because of the availability heuristic, people are far more fearful of some infectious disease (say, ebola) than statistical reality warrants, and that a nudge gives them a more accurate sense of the risk of getting the disease. Many nudges promote people's capacity for agency; they do not undermine it. A nudge may promote autonomy if it counteracts unrealistic optimism or allows people to see aspects of situations that would otherwise be shrouded or invisible. As we have seen, some people are troubled by the potential tension between government nudges and individual agency. But for many such nudges, there is no such tension; because of nudges, agency is actually increased.

It is also important to see that autonomy does not require choices everywhere. It does not justify an insistence on active choosing in all contexts. If we had to make choices about everything that affects us, we would quickly be overwhelmed. There is a close relationship between time-management and autonomy. People should be allowed to devote attention to the questions that, in their view, deserve their attention. If people have to make choices everywhere, their autonomy is reduced, if only because they cannot focus on those activities that seem to them most worthy of their attention.¹¹⁶

It is nonetheless true that on grounds of autonomy (as well as welfare), the best choice architecture often calls for active choosing. Even though default rules preserve freedom of choice, they might intrude on autonomy, at least if they do not track people's likely choices. The problem is that because of the force of inertia, people might not reject harmful defaults.¹¹⁷ If so, there is arguably an intrusion on people's autonomy because they will end up with outcomes that they did not specifically select. Consider, for example, a default rule that says that if you do not indicate otherwise, you are presumed to be a member of a political party that you do not like, or to want your estate to go to the Vatican, or prefer your organs to go to people with a certain skin color. Even though people can opt out, default rules can intrude on autonomy insofar as they impose that burden on people—and insofar as the particular rules a) might stick because of that very burden and b) do not reflect what informed people would like.

114. See George Loewenstein et al., *supra* note 74.

115. See CONLY, *supra* note 64, at 36 ("Even if we accept that individuals have rights, and thus claims not to be harmed by others in certain ways, or to have (yet) others defend them in these claims, why would there be such a right here, where the point of the action is to help the person achieve what in the long run, he wants, and what he would want not if he were not a flawed thinker?").

116. See MULLAINATHAN & SHAFIR, *supra* note 23, at 41.

117. See REBONATO, *supra* note 1, at 82-83.

To return to one of my general themes: Whether the interest in autonomy calls for active choosing, as opposed to reliance on a default rule, depends on the circumstances. Along some dimensions, default rules are actually superior to active choosing on autonomy grounds. If people choose not to choose, or if they would make that choice if asked, it is an insult to their autonomy to force them to choose.¹¹⁸ And if people would like to choose, a default rule does not deprive them of that choice; they can reject the default. Even in the face of inertia, many people will do so.¹¹⁹ Preservation of freedom of choice is not sufficient, but it goes some distance toward ensuring that people's autonomy is respected. The same can be said for a requirement that any paternalistic nudges focus on people's own ends and otherwise have legitimate goals. But with respect to autonomy, a continuing problem lies in the possibility of manipulation; I will turn to that problem below.

Insofar as nudges respect freedom of choice, they are, along an important dimension, preserving people's autonomy. Once choice architects coerce people, they are no longer merely nudging.¹²⁰ But if they are focused on autonomy, skeptics might again emphasize that the problem of coercion cannot be entirely avoided with some nudges. We have seen that because of the power of inertia, people might accept (passively) a default rule even though they have no enthusiasm for the outcome that it produces, and would reject that outcome if they focused on the issue involved.¹²¹

We might doubt whether such situations are properly described as involving coercion. No one is being forced to do anything. But there is certainly a risk that a default rule will produce harmful results. Choice architects need to take account of that risk and take steps to ensure that freedom of choice is maintained and real. So long as it is, coercion is not involved.

C. Dignity

The idea of "dignity" is complex and contested.¹²² On one view, autonomy and dignity are closely related.¹²³ But in both politics and law, dignity sometimes occupies a distinctive place. In German constitutional law, for example, dignity has a prominent role, with the Constitution starting with these words: "Human dignity is inviolable [*unantastbar*]. To respect it and protect it is the duty of all state power."¹²⁴

118. See Sunstein, *supra* note 17, at 22.

119. See *id.* at 30.

120. Recall that an initiative might require its objects (employers, advertisers) to nudge others (employees, consumers); consider a mandatory fuel economy label.

121. For an excellent discussion, see REBONATO, *supra* note 1, at 205-11.

122. For an especially valuable discussion, see MICHAEL ROSEN, *DIGNITY: ITS HISTORY AND MEANING* (2012). In the legal context, see Christopher McCrudden, *Human Dignity and Judicial Interpretation of Human Rights*, 19 *EURO. J. INT'L L.* 655 (2008).

123. This view is described but not endorsed in ROSEN, *supra* note 122, at 4-5.

124. *Id.* at 78.

We might begin by suggesting that the antonym of autonomy is coercion; the antonym of dignity is humiliation.¹²⁵ As I understand it here, the idea of dignity requires respect for people as agents; a person's dignity is not respected if he is treated as an infant or a thing, subject to the superior authority of another. A noncoercive intervention, respectful of people's autonomy, might nonetheless compromise dignity. Imagine, for example, that an employer allows employees a great deal of room for free choice but speaks to them as if they are children, who have difficulty in engaging in even minimal forms of reasoning. A police officer can also fail to treat citizens with dignity even if he respects their autonomy (though again, we could understand autonomy in a way that is coextensive with dignity).

Some nudges might seem to compromise dignity and respect for persons. They might humiliate people. As we shall see, this objection is both interesting and important, especially when it is combined with a concern about manipulation.¹²⁶ Imaginable forms of choice architecture could indeed undermine dignity.

There are of course large questions about the place of dignity in ethics.¹²⁷ On one (admittedly unconventional) view, dignity is properly part of an assessment of welfare. If people feel humiliated, or think that they have been treated disrespectfully, they suffer a welfare loss. That loss might be extremely serious. In any assessment of welfare consequences, such a loss must be considered. It might turn out to be exceedingly important, and to argue against particular nudges.

To avoid obtuseness, a good welfarist should acknowledge that an offense to dignity is qualitatively distinct.¹²⁸ In its very nature, it is a different kind of loss from the loss of, say, money, or an opportunity to visit a beach. But on the welfarist view, a dignity loss is just one kind of welfare loss, to be weighed against the other goods that are at stake. Suppose, for purposes of argument,¹²⁹ that a graphic and highly emotional appeal, triggering strong, System 1 emotions in order to discourage people from smoking, is plausibly seen as an offense to dignity—as a way of treating smokers disrespectfully (and perhaps infantilizing them). Some smokers might so regard such an appeal and object for that reason. But a welfarist might be willing to support the emotional appeal, notwithstanding the relevant loss, if it saves a significant number of lives.

125. See AVISHAI MARGALIT, *THE DECENT SOCIETY* 9-27 (1998).

126. See Waldron, *supra* note 110. On paternalism and respect, see generally Cornell, *supra* note 106 (offering an account of the distinctive wrongs of paternalism).

127. See MARGALIT, *supra* note 125, at 9-27; see generally ROSEN, *supra* note 122; Charles Beitz, *Human Dignity in the Theory of Human Rights*, 41 PHIL. & PUB. AFF. 259 (2013); Thomas Christiano, *Two Conceptions of Human Dignity as Persons* (2008) (unpublished manuscript).

128. On qualitative distinctions and their importance, see generally ELIZABETH ANDERSON, *VALUE IN ETHICS AND ECONOMICS* (1993).

129. For discussion, see Kyle Rozema, *Economic Theory Lost in Translation: Will Behavioral Economics Reshape the Compelled Commercial Speech Doctrine?* 23 CORNELL J.L. & PUB. POL'Y 85, 90 (2013).

On a different view, an insult to dignity is not merely part of a welfarist calculus. Such an insult does not depend on people's subjective feelings, and it is a grave act, perhaps especially if it comes from government. An insult to dignity, understood as a failure to treat people with respect, should not be permitted unless it has an overwhelmingly strong justification. If we endorse this view, it is especially important to ask whether nudges offend human dignity.

To return to my general plea: The force of the objection depends on the particular nudge. A GPS insults no one's dignity. Disclosure of factual information can hardly be seen as an offense to dignity—certainly if the information is useful and not based on a false and demeaning belief that people need it. But we can easily imagine nudges that would offend one or another conception of dignity. Consider a public health campaign, directed at the prevention of obesity, that stigmatized and humiliated people who are overweight by portraying them in a demeaning light.¹³⁰ Or consider, as a somewhat more difficult case, an antismoking campaign that did the same for smokers. Here again, the fact that nudges preserve freedom of choice, and do not require anyone to do anything, should not be taken as a kind of license to do anything at all. It is possible to imagine public education campaigns that offend dignity, though admittedly the more familiar real-world campaigns do not have anything approaching that vice.

It might also count as an insult to dignity, and a form of infantilization, if the government constantly reminds people of things that they already know. Every child, and everyone who was once a child, can recall this form of infantilization, and it is not always absent from adult life as well. If people are informed of the same thing every hour or even every day (say, by their spouse, by their doctor, or by some public official), they might legitimately feel that their dignity is not being respected.

At the same time, most reminders, warnings, and uses of social norms do not demean anyone.¹³¹ Nor are choice architects likely to be humiliating choosers if they spread information about what most people do.¹³² In some cases, however, the concern about dignity might become more serious. If people are frequently reminded that a due date is coming, they might feel as if they are being treated like children. Warnings can run into the same concern insofar as they are repetitive or condescending, or are meant to trigger strong emotions instead of merely giving people a sense of factual realities.¹³³

Here as well, there is no objection to the relevant nudges in the abstract, but there could be concerns about specific types of nudges. At the same time, it must be emphasized that the relevant offense to dignity—coming from unwelcome

130. I am grateful to Gertrude Lubbe-Wolff for this example.

131. On the functions of norms, see EDNA ULLMANN-MARGALIT, *THE EMERGENCE OF NORMS* (1976).

132. See Allcott, *supra* note 66, at 1084.

133. See *id.* at 1083.

and numerous reminders or warnings—is relatively minor, and from the standpoint of the concerns that have produced the focus on dignity in the Western political tradition, it is laughably modest.

What is the relationship between dignity and default rules? If an employer automatically enrolls employees into retirement and health care plans, dignity is hardly at risk. If a cell phone company adopts a series of default rules for the phone and the contractual arrangement, nothing need be amiss in terms of dignity. But we could imagine harder (if pretty fanciful) cases. Suppose that the government insisted on “default meals” in various restaurants so that people would be given certain healthy choices unless they specifically chose otherwise. Put to one side the fact that with respect to restaurants, this approach is a mandate, not a mere nudge. A plausible response is based on a concern about dignity: Why shouldn’t free people be asked to select what they want? Or suppose that a government specified a “default exercise plan” for adults, so that they would be presumed to want to engage in certain activities unless they opted out. People might offer the same plausible response, perhaps with considerable agitation.

Note that default rules of this kind might be objectionable for both welfarists and nonwelfarists. Welfarists might want to focus on people’s subjective feelings. If people believe that they are being treated as children, and if they object to that treatment, welfarists should count this objection in their assessment of costs and benefits. Nonwelfarists would insist that the offense to dignity is objectionable even if it has some kind of welfarist justification.

In extreme situations, default rules could indeed be a serious affront to dignity. If so, there should be a strong presumption against them (whatever our foundational commitments).¹³⁴ But it would be a mistake to use extreme situations or barely imaginable cases as a reason to challenge default rules in general. People are not treated disrespectfully if an institution adopts a double-sided default for printing or if they are automatically enrolled in health insurance or retirement plans. The objection from dignity has far more force in the abstract than in the context of all, or nearly all, real-world cases in which default rules are actually at work.

D. Manipulation

I have referred at several points to the problem of manipulation, which is a possible objection to various nudges. If manipulation is wrong, it is likely because it reduces welfare or intrudes on autonomy or dignity. But the topic of manipulation is both interesting and important, and it deserves separate treatment.

134. Perhaps the presumption could be rebutted with a sufficiently strong consequentialist justification—as, for example, by showing that many lives would be saved with the appropriate default.

1. *Definitions.* To deal with this objection, we need to say something about the complex idea of “manipulation.”¹³⁵ It should be clear that an action does not count as manipulative merely because it is an effort to persuade or warn people or to alter their behavior. If you warn a driver that he is about to get into a crash, you are not engaged in manipulation. The same is true if you remind someone that a bill is due. A calorie label and an energy efficiency label are not ordinarily counted as forms of manipulation.¹³⁶ Economic incentives or penalties can have large effects on behavior, but they are hardly manipulative.

It is not clear that the idea of “manipulation” is a unitary concept, or that we can identify necessary and sufficient conditions.¹³⁷ No one should doubt that manipulation takes various forms, even if they are only loosely identified with one another.¹³⁸ An organizing idea is that when one is being manipulated, one is being treated as a kind of “puppet on a string.”¹³⁹ No one wants to be someone’s puppet, and it is especially bad to be a puppet of government. But T. M. Wilkinson is correct to say that “one should not rush to judgement when trying to decide whether nudging manipulates.”¹⁴⁰

Consider one definition: *An action counts as manipulative if it attempts to influence people in a way that does not sufficiently engage or appeal to their capacities for reflective and deliberative choice.* This definition captures the essentials of several different accounts. For example, Wilkinson contends that manipulation “is a kind of influence that bypasses or subverts the target’s rational capacities.”¹⁴¹ Making an explicit link to autonomy, Wilkinson urges that manipulation “subverts and insults a person’s autonomous decision making,” in a way that treats its objects as “tools and fools.”¹⁴² He contends that

135. For a helpful discussion, see Anne Barnhill, *What is Manipulation?*, in *MANIPULATION: THEORY AND PRACTICE* 50, 72 (Christian Coons & Michael Weber eds. 2014).

136. A qualification is necessary. If a disclosure requirement focuses on one of many aspects of a situation, and fixes people’s attention on that aspect, a charge of manipulation would not be unreasonable. Consider the controversy over the idea that sellers should have to disclose that food has genetically modified organisms (GMOs). See Charles Noussair et al., *Do Consumers Really Refuse to Buy Genetically Modified Food?*, 114 *ECON. J.* 102 (2004). For those who object to compulsory labeling about GMOs, there is a plausible claim that labels are a form of manipulation, activating public concern where there is no objective reason for that concern.

137. For a number of valuable treatments, see *MANIPULATION: THEORY AND PRACTICE* (Christian Coons & Michael Webster eds., 2014).

138. A valuable cautionary note: “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.” *Id.* at 344.

139. See T. M. Wilkinson, *Nudging and Manipulation*, 61 *POL. STUDIES* 341, 342 (2013).

140. See *id.*

141. Christian Coons & Michael Webster, *Introduction*, in *MANIPULATION: THEORY AND PRACTICE*, *supra* note 137, at 11.

142. Wilkinson, *supra* note 139, at 345.

“manipulation is intentionally and successfully influencing someone using methods that pervert choice.”¹⁴³

In a similar account, Ruth Faden and Tom Beauchamp define psychological manipulation as “any intentional act that successfully influences a person to belief or behavior by causing changes in mental processes *other than those involved in understanding*.”¹⁴⁴ Joseph Raz suggests that “[m]anipulation, unlike coercion, does not interfere with a person’s options. Instead it perverts the way that person reaches decisions, forms preferences or adopts goals.”¹⁴⁵ Of course the idea of “perverting” choice, or people’s way of reaching decisions or forming preferences, is not self-defining; it is often taken to refer to methods that do not appeal to, or produce, conscious deliberation. If so, the objection to manipulation is that it “infringes upon the autonomy of the victim by subverting and insulting their decision-making powers.”¹⁴⁶

The objection certainly applies to lies, which attempt to alter behavior not by engaging people on the merits and asking them to decide accordingly, but by enlisting falsehoods in the service of the liar’s goals. A lie is disrespectful to its objects, not least if it attempts to exert influence without asking people to make a deliberate choice in light of relevant facts. In harder cases, the challenge is to concretize the ideas of “subverting” and “insulting.”

Subliminal advertising should also be deemed manipulative and insulting, because it operates “behind the back” of the person involved without appealing to his conscious awareness. People’s decisions are affected in a way that bypasses their own deliberative capacities. If this is the defining problem with subliminal advertising, we can understand why involuntary hypnosis would also count as manipulative. But almost no one favors subliminal advertising, and to say the least, the idea of involuntary hypnosis does not have much appeal. The question is whether admittedly taboo practices can shed light on interventions, including nudges, which might be able to command broader support.

2. *Are nudges manipulative?* Some forms of framing count as nudges, and they could plausibly be characterized as manipulative. Suppose that public officials try to persuade people to engage in certain behavior with the help of relative risk information: “If you do not do X, your chances of death from heart disease will triple!”¹⁴⁷ Suppose that for the relevant population, the chance of death from heart disease is very small—say, one in 100,000—and that people are far more influenced by the idea of “tripling the risk” than they would be if they learned that they could increase a 1/100,000 risk to a 3/100,000 risk. On one

143. See *id.* at 347. An especially valuable and subtle account, broadly compatible with that offered here but adding illuminating refinements, is Barnhill, *supra* note 135, at 50-72.

144. RUTH FADEN & TOM BEAUCHAMP, A HISTORY AND THEORY OF INFORMED CONSENT, 354-68 (1986).

145. RAZ, *supra* note 83, at 377.

146. See Wilkinson, *supra* note 139, at 351.

147. See *id.* at 347, which also uses this example.

view, the choice of the relative risk frame does not sufficiently respect people's decisionmaking capacities—and it appeals directly to System 1. If it is manipulative, the problem is its failure sufficiently to engage people's reflective or deliberative capacities.

Or suppose that choice architects are alert to the power of loss aversion,¹⁴⁸ and hence they nudge people by using the “loss frame,” so as to trigger people's concern about the risks associated with obesity and excessive energy consumption. They might deliberately choose to emphasize, in some kind of information campaign, how much people would lose from not using energy conservative techniques, rather than how much people would gain from using such techniques. If the use of loss aversion can be counted as manipulative, it is because it does not sufficiently appeal to people's deliberative processes but instead tries to trigger the negative feelings that are associated with losses.

But we have to be careful here, because an intelligible understanding of manipulation might sweep up and perhaps condemn a great deal of conduct that is generally seen as unobjectionable, and reasonably so. It would be fussy, and far too stringent, to condemn all such conduct on ethical grounds, even if the word “manipulation” is plausibly applied to it. Some actions are unquestionably manipulative, and others fall within the periphery of the concept, and an ethical evaluation must depend on the context and on the roles of the relevant choice architects.

Much of modern advertising is directed at System 1, with attractive people, bold colors, and distinctive aesthetics. (Consider advertisements for Viagra.) Cell phone companies, restaurants, and clothing stores use music and colors in a way that is designed to “frame” products in a distinctive manner. Doctors, friends, and family members (including spouses) often do something quite similar. Is romance an exercise in manipulation? Maybe so.¹⁴⁹ Is medical care? Is the use of social media? A great deal of conduct, however familiar, can be counted as manipulative in some relevant sense, but it would be extreme to condemn it for that reason.

Sarah Conly contends that when nudges are at work:

Rather than regarding people as generally capable of making good choices, we outmaneuver them by appealing to their irrationality, just in more fruitful ways. We concede that people can't generally make good decisions when left to their own devices, and this runs against the basic premise of liberalism, which is that we are basically rational, prudent creatures who may thus, and should thus, direct themselves autonomously.¹⁵⁰

148. See EYAL ZAMIR, *LAW, PSYCHOLOGY, AND MORALITY: THE ROLE OF LOSS AVERSION* 204 (2014).

149. I am aware of no detailed treatment of this question, but for relevant discussion, see Eric Cave, *Unsavoury Seduction and Manipulation*, in *MANIPULATION: THEORY AND PRACTICE*, *supra* note 137, at 176.

150. CONLY, *supra* note 64, at 30.

Only is rightly linking a concern about manipulation with the ideas of autonomy and dignity. But she is making a strong charge, one that is not fairly leveled against most kinds of nudges. Recall that many nudges are educative, and they do not enlist or exploit System 1 or attempt in any way to “outmaneuver” people. But consider some testing cases of real-world nudging, where the charge of manipulation is not self-evidently misplaced:

(a) Choice architects might choose a graphic health warning, on the theory that an emotional, even visceral presentation might have significant effects.¹⁵¹

(b) They might be aware that a statement that a product is “90 percent fat-free” has a different impact from a statement that a product is “10 percent fat,” and they might choose the frame that has the desired effect.

(c) They might make a strategic decision about how to present social norms, knowing that the right presentation—for example, emphasizing behavior within the local community—could have a large impact on people’s behavior.

(d) They might decide to list options—in a cafeteria or on a form—so as to make it more likely that people will make certain choices.

It is an understatement to say that none of these cases involves the most egregious forms of manipulation. There is no lying and no deceit. But is there an effort to subvert or to insult people’s decision-making powers? It is not absurd to say that at least some of these cases, the answer is yes. That answer does not resolve the ethical issues. In these cases, any manipulation seems modest, and in all of them, I believe that the relevant choice architecture can be justified. But it is not illegitimate to raise concerns.

3. *Transparency and manipulation.* I have said that government should be transparent about what it is doing. It should not hide its actions or its reasons for those actions. Does transparency rebut the charge of manipulation? Probably not. If government engages in egregious forms of manipulation, transparency is not a defense.¹⁵² A genuine insult to autonomy and dignity, in the form of a subversion of people’s decisionmaking capacities, does not become acceptable merely because people are allowed to know about it.

We could easily imagine cases where full democratic control, alongside a high degree of transparency, is plainly insufficient to rebut a charge of manipulation. Imagine that a democratic government adopted, freely and openly, a program of subliminal advertising—designed, let us say, for purposes of promoting public health (such as reducing smoking) rather than for illicit

151. See Christine Jolls, *Product Warnings, Debiasing, and Free Speech: The Case of Tobacco Regulation*, 169 J. INST. AND THEORETICAL ECON. 53 (2013).

152. On the relationship between consent and manipulation, see Wilkinson, *supra* note 139, which suggests that consent can provide justification.

purposes. The program would be objectionable on the ground that it would be unacceptably manipulative, even if it was adopted openly and would produce good consequences.¹⁵³

4. *Consent and justification, core and periphery.* Perhaps a graphic health warning could be counted as manipulative if it is designed to target people's emotions, rather than to inform them of facts.¹⁵⁴ But what if the warning is explained, in public, on exactly that ground? What if a warning is justified as effective because it appeals to people's emotions, and thus saves lives? What if it is welcomed by the relevant population – say, the public at large, or smokers – for exactly that reason?¹⁵⁵ And what if a graphic warning ends up giving people a more accurate sense of the risks that they actually face? Similar questions might be asked about strategic uses of framing effects, social norms, and order effects. T. M. Wilkinson convincingly argues that it is too crude to say that manipulation infringes upon autonomy, because “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.”¹⁵⁶

We could understand consent as suggesting support from System 2, which might welcome a little manipulation (or possibly a great deal) as a way of cabining the adverse effects of System 1 (recall present bias). To be sure, there are dangers in authorizing public officials to pursue this line of argument. Individual consent is one thing; collective consent through majorities is another. Perhaps majorities, approving of such measures, should not be allowed to authorize measures that manipulate minorities, who oppose them. But in certain contexts, the argument is more than plausible. Imagine, for example, a public education campaign that is designed to reduce the risks associated with texting while driving, or an effort to combat the use of drugs or to convince people to stay in school. Many such campaigns are vivid and have an emotional component; they can be understood as efforts to combat self-control problems and to focus people on the long term.

If government is targeting System 1—perhaps through framing or emotionally evocative appeals—it may be responding to the fact that System 1 has already been targeted, and to people's detriment. In the context of cigarettes, for example, it is plausible to say that previous manipulations—including advertising and social norms—have influenced people to become smokers. If this

153. I do not mean to resolve here the question whether good consequences could justify an admittedly manipulative action. In ordinary life, we could imagine a manipulative act—designed, say, to lead a child, a spouse, or a parent to take medicine—that would have an adequate justification.

154. There are also possible First Amendment issues. Is it unconstitutional to require companies to include graphic warnings about the harms associated with their own products if the requirement has a behavioral motivation and is understood to be targeting System 1?

155. Note the overwhelming public approval of graphic warnings for smokers, found in Sunstein, *Do People Like Nudges?*, *supra* note 78.

156. Wilkinson, *supra* note 139, at 345.

is so, perhaps we can say that public officials are permitted to meet fire with fire. But some people might insist that two wrongs do not make a right—and that if the government seeks to lead people to quit, it must treat them as adults by appealing to their deliberative capacities. Reasonable people can disagree about this question; my own view is that framing and emotionally evocative appeals are legitimate tools in this context, especially if significant numbers of lives are at stake.

Nonetheless, it is reasonable to worry that, even with full transparency, some manipulation may be involved whenever a choice architect is targeting emotions or seeking a formulation that will be effective because of how it interacts with people's intuitive or automatic thinking. But there are degrees of manipulation, and there is a large difference between a lie and an effort to frame an alternative in an appealing light. In ordinary life, we would not be likely to accuse our friends or loved ones of manipulation if they characterized one approach as favored by most members of our peer group, or if they emphasized the losses that might accompany an alternative that they abhor, or if they accompanied a description of one option with a frown and another with a smile.

Actions that are plausibly characterized as manipulative fall along a continuum, and if a doctor or a lawyer uses body language to support or undermine one or another alternative, it would be pretty fussy to raise objections about "subverting" or "perverting" the deliberative processes of a patient or client. No one should deny that at least some nudges can be considered as manipulative within ordinary understandings of that term. I have emphasized that any action by government, including nudging, must meet a burden of justification. But when nudges fall outside the core and only within the periphery of the concept of manipulation, when they have legitimate purposes, when they would be effective, and when they do not diverge from the kinds of influences that are common and unobjectionable in ordinary life, the burden of justification is generally met.

We have seen most nudges are not manipulative in any relevant sense. But to the extent that some of them are, we can imagine a reasonable objection or concern, whose force depends on the degree of the manipulation. We might well favor an absolute or near-absolute taboo on lying or deception on government's part, for welfarist or nonwelfarist reasons.¹⁵⁷ But for reasons discussed above, we should be more lenient toward emotional appeals and framing. Because government always faces a burden of justification, one question is whether such approaches produce significant welfare gains. If a graphic health warning saves many lives, is it unacceptable if and because it can be counted as a mild form of manipulation? That is most doubtful. A welfarist would want to make an all-things-considered judgment about the welfare consequences.

157. Under true emergency conditions—for example, when national security is genuinely threatened—it is possible that this constraint can be overcome.

It is true that some people, focused on autonomy as an independent good, would erect a strong and perhaps conclusive presumption against clear or egregious cases of manipulation.¹⁵⁸ But, at least in general, the modest practices discussed here strain the boundaries of the concept, and it would be odd to rule them off-limits.

E. Biased Officials

Choice architects are emphatically human, and fully subject to behavioral biases; they are often unreliable. The growing field of *behavioral public choice* draws on this point to offer an account of official error.¹⁵⁹ It is reasonable to object to some nudges, and to some efforts to intervene in existing choice architecture, on the ground that the choice architects might blunder.¹⁶⁰ They might lack important information (the knowledge problem). They might be biased, perhaps because their own parochial interests are at stake (the public choice problem). They might themselves display behavioral biases—such as present bias, optimistic bias, or probability neglect. In a democratic society, public officials are responsive to public opinion, and if the public is mistaken, officials might be mistaken as well.

It is unclear whether and to what extent this objection is a distinctly ethical one, but it does identify an important cautionary note. One reason for nudges, as opposed to mandates and bans, is that choice architects may err.¹⁶¹ No one should deny that proposition, which argues strongly in favor of choice-preserving approaches. If choice architects blunder, at least it can be said that people are entitled to go their own way. And if we emphasize the risk of official error, we might want to avoid nudges and choice architecture as well.

The initial response to this objection should be familiar: Choice architecture is inevitable. When choice architects act, they alter the architecture; they do not create an architecture where it did not exist before. A certain degree of nudging from the public sector cannot be avoided, and there is no use in wishing it away. Nonetheless, choice architects who work for government might decide that it is best to try to rely on free markets and to trust free markets and invisible-hand mechanisms. If so, they would select (or accept) choice architecture that reflects those mechanisms.

This idea raises many conceptual and empirical puzzles, to which I have referred above, and which I will not engage in detail here. The question is whether it is so abstract, and so rooted in dogmas, that it ought not to command

158. Cf. WHITE, *supra* note 32.

159. For one example, see Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683 (1999). For an overview, see Jan Schnellenbach & Christian Schubert, *Behavioral Public Choice: A Survey* (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2390290.

160. REBONATO, *supra* note 1, at 221-226.

161. See Sunstein, *supra* note 105, at 211.

support. To be sure, free markets have many virtues. But in some cases, disclosure, warnings, and reminders can do far more good than harm.¹⁶² As we have seen, active choosing is sometimes inferior to default rules. Someone has to decide in favor of one or another, and in some cases, that someone is inevitably the government. It is true that distrust of public officials will argue against nudging, at least where it is avoidable, but if it is dogmatic and generalized, such distrust will likely produce serious losses in terms of welfare, autonomy, and even dignity. Invisible hands can and do compromise all of those things.¹⁶³

Conclusion

It is pointless to object to nudges and choice architecture as such. Human beings cannot live in a world without them. Spontaneous orders and invisible hands have many virtues, but they nudge. Even the most minimal government must create choice architecture of many different kinds. The common law itself influences people's decisions whether or not it was designed to do so (or designed at all). Consider the effects of default rules, of the sort that are pervasive in the law of property, contract, and tort.

Many nudges, and many changes in choice architecture, are not merely permissible on ethical grounds; they are actually required. On grounds of welfare, the point should be straightforward; much nudging promises to increase social welfare. But the point holds for autonomy and dignity as well. In some cases, there is an imaginable tension between nudges and human agency, whose importance has everything to do with welfare, autonomy, and dignity. But whether nudges compromise agency depends on their content, not their status as nudges. As we have seen, many nudges promote agency.

It is important to have a sufficiently capacious sense of the category of nudges, and a full appreciation of the differences among them. To make progress, ethical concerns must be brought into close contact with particular examples. It is true that any alterations in choice architecture, including those that preserve freedom, can run into serious and even convincing ethical objections—most obviously, where the underlying goals are illicit. But where the goals are legitimate, nudges are less likely to run afoul of ethical constraints, not least when they promote informed choices (as in the case of reminders). Transparency and public scrutiny are important safeguards, especially when public officials are responsible for nudges and choice architecture. Nothing should be hidden or covert.

The history of freedom-respecting nations is full of changes in choice architecture that have been motivated by an aspiration to realize the highest and most enduring of national ideals. In moving closer to those ideals, new nudges, and new forms of choice architecture, will prove indispensable.

162. See BAR-GILL, *supra* note 93, at 32.

163. See AKERLOF & SHILLER, *supra* note 22, at 2-8.