

ISSN 1936-5349 (print)  
ISSN 1936-5357 (online)

# HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

## THE ETHICS OF NUDGING

Cass R. Sunstein

Discussion Paper No. 806

01/2015

Harvard Law School  
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:  
[http://www.law.harvard.edu/programs/olin\\_center/](http://www.law.harvard.edu/programs/olin_center/)

The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=2526341>

## **The Ethics of Nudging**

Cass R. Sunstein\*

### **Abstract**

*This essay defends the following propositions. (1) It is pointless to object to choice architecture or nudging as such. Choice architecture cannot be avoided. Nature itself nudges; so does the weather; so do spontaneous orders and invisible hands. The private sector inevitably nudges, as does the government. It is reasonable to object to particular nudges, but not to nudging in general. (2) In this context, ethical abstractions (for example, about autonomy, dignity, and manipulation) can create serious confusion. To make progress, those abstractions must be brought into contact with concrete practices. Nudging and choice architecture take diverse forms, and the force of an ethical objection depends on the specific form. (3) If welfare is our guide, much nudging is actually required on ethical grounds. (4) If autonomy is our guide, much nudging is also required on ethical grounds. (5) Choice architecture should not, and need not, compromise either dignity or self-government, though imaginable forms could do both. (6) Some nudges are objectionable because the choice architect has illicit ends. When the ends are legitimate, and when nudges are fully transparent and subject to public scrutiny, a convincing ethical objection is less likely to be available. (7) There is, however, room for ethical objections in the case of well-motivated but manipulative interventions, certainly if people have not consented to them; such nudges can undermine autonomy and dignity. It follows that both the concept and the practice of manipulation deserve careful attention. The concept of manipulation has a core and a periphery; some interventions fit within the core, others within the periphery, and others outside of both.*

### **I. Introduction**

---

\* Robert Walmsley University Professor, Harvard University. This essay is meant as the basis for remarks on the ethics of choice architecture and nudging, to be delivered at a conference on that topic at Humboldt University in Berlin in January 2015. I am grateful to Matthew Lipka, Martha Nussbaum, Lucia Reisch, and Adrian Vermeule for valuable comments on a previous draft. Continuing thanks for Richard Thaler for joint work and innumerable discussions of these issues; he should not be held responsible for my errors here.

1. Nudges are interventions that steer people in particular directions but that also allow them to go their own way.<sup>1</sup> 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.<sup>2</sup> 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 is not a nudge.
2. When people make decisions, they do so against a background consisting of choice architecture.<sup>3</sup> 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 (such as leaving without making a purchase). Even if the layout of a department store is a result of chance, or does not reflect any effort to steer people, it will have consequences for what people select.<sup>4</sup> If people see certain items first, they are more likely to buy them.<sup>5</sup>
3. Attention is a scarce resource. When applications (for loans, for educational opportunities, for training, for financial benefits of any kind) are complex, people may not apply; a great deal of money might be lost as a result.<sup>6</sup> Spontaneous orders nudge no less than intentional designs, and invisible hands can nudge every bit as much as the most visible ones. To be sure, spontaneous orders and invisible hands may be less dangerous than intentional designs, but they are nonetheless forms of choice architecture.
4. For the future, we could imagine new forms of choice architecture that are designed to improve antipoverty programs<sup>7</sup>; environmental programs<sup>8</sup>; energy programs<sup>9</sup>; retirement and social security programs<sup>10</sup>;

---

<sup>1</sup> See Richard Thaler and Cass R. Sunstein, *Nudge* (2008).

<sup>2</sup> There are some complexities here. For example, a nudge might impose psychic costs (as in the case of graphic health warnings), and it also might create inconvenience costs (as in the placement of goods in a cafeteria). On these issues, see Brian Wansink, *Slim By Design* (2014); Cass R. Sunstein, *Why Nudge?* (2014).

<sup>3</sup> See *id.*

<sup>4</sup> See Brian Wansink, *Slim By Design* (2014).

<sup>5</sup> Eran Dayan and Maya Bar-Hillel, *Nudge to Nobesity II: Menu Positions Influence Food Orders*, 6 *Judgment and Decision Making* 333 (2011).

<sup>6</sup> See Benjamin Keys et al., *Failure to Refinance* (2014), available at <http://www.nber.org/papers/w20401>

<sup>7</sup> See Sendhil Mullainathan and Eldar Shafir, *Scarcity* (2013).

<sup>8</sup> Cass R. Sunstein and Lucia Reisch, *Automatically Green*, 38 *Harv Env L Rev* 128 (2014).

anti-obesity programs<sup>11</sup>; educational programs<sup>12</sup>; health care programs; and programs to increase organ donation.<sup>13</sup> We could also imagine forms of choice architecture that are designed to combat race and sex discrimination,<sup>14</sup> to help disabled people, and to promote economic growth. A great deal of future work needs to be devoted to choice architecture in these and related domains.

5. Many nudges are designed to help choosers, by increasing their welfare or promoting their autonomy.<sup>15</sup> But some nudges are designed to help third parties – for example, by reducing environmental harm.<sup>16</sup> When third-party effects are involved, or when the goal is to solve a collective action problem, there is a standard argument on behalf of government action. Nudges might be used instead of mandates and bans, or they might be complementary – as, for example, when employers automatically enroll employees into a health insurance plan, in a system in which health insurance is mandatory.
6. No one should doubt that certain nudges, and certain kinds of choice architecture, can raise serious ethical questions.<sup>17</sup> Consider, for example, a government that used nudges to promote discrimination on the basis of race, sex, or religion. Even truthful information (for example, about crime rates) might fan the flames of violence and prejudice. Groups or nations that are committed to violence often enlist nudges in their cause. Even when nudges do not have illicit ends, it is possible to wonder whether those who enlist them are treating people with respect.

---

<sup>9</sup> Id.

<sup>10</sup> Ryan Bubb and Richard Pildes, Why Behavioral Economics Trims its Sails, Harv L Rev (2014).

<sup>11</sup> See Brian Wansink, Slim By Design (2014).

<sup>12</sup> See Adam Lavecchia et al., Behavioral Economics of Education (2014), available at <http://www.nber.org/papers/w20609>

<sup>13</sup> For an interesting empirical result, see Judd Kessler and Alvin Roth, Don't Take 'No' For An Answer: An Experiment With Actual Organ Donor Registrations (2014), available at <http://www.nber.org/papers/w20378> (finding that required active choosing has a smaller effect, in terms of getting people to sign up for organ donation, than prompted choice).

<sup>14</sup> See Iris Bohnet et al., When Performance Trumps Gender Bias: Joint Versus Separate Evaluation (2013), available at <http://www.montana.edu/nsfadvance/documents/PDFs/resources/WhenPerformanceTrumpsGenderBias.pdf>

<sup>15</sup> See Oren Bar-Gill, Seduction By Contract (2012).

<sup>16</sup> See Sunstein and Reisch, supra note.

<sup>17</sup> See Mark White, The Manipulation of Choice: Ethics and Libertarian Paternalism (2013); Riccardo Rebonato, Taking Liberties (2011).

7. Possible concerns about nudging and choice architecture point to four foundational commitments: (1) welfare, (2) autonomy, (3) dignity, and (4) self-government. 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 money<sup>18</sup>; an unhelpful default rule could fall in 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.<sup>19</sup> Excessive pollution is, in part, a product of unhelpful choice architecture.<sup>20</sup>
8. 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 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?’”<sup>21</sup> 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.<sup>22</sup> Human beings cannot live without some kind of weather. Nature nudges.
9. We can imagine the following view: Choice architecture is unavoidable, to be sure, but it might be the product of nature or some kind of spontaneous order, rather than of conscious design, or of the action of any designer. Invisible hand mechanisms<sup>23</sup> often produce choice architecture. Alternatively, choice architecture might be the product of a genuinely random process (and a choice architect might intentionally opt for randomness, on the ground that it has a kind of neutrality<sup>24</sup>).
10. On certain assumptions, self-conscious choice architecture is especially dangerous, because it is explicitly directed at achieving certain goals. But what are those assumptions, and are they likely to be true? Why and when would spontaneous order be benign? (Is there some kind of social

---

<sup>18</sup> See the weak effects of the opt-out designed in Robert Letzler, *Knowing When To Quit: Default Choices, Demographics, and Fraud* (2014), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2512276](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512276)

<sup>19</sup> For relevant discussion, see Sunstein and Reisch, *supra* note.

<sup>20</sup> See *id.*

<sup>21</sup> Available at <http://moreintelligentlife.com/story/david-foster-wallace-in-his-own-words>

<sup>22</sup> Meghan R. Busse et al., *Projection Bias in the Car and Housing Markets* (2014), available at <http://www.nber.org/papers/w18212>

<sup>23</sup> For a superb discussion, see Edna Ullmann-Margalit, *The Invisible Hand and the Cunning of Reason*, 64 *Social Research* 181 (1997).

<sup>24</sup> See Jon Elster, *Solomonic Judgments* (1989).

Darwinism here<sup>25</sup>?) What is so good about randomness? We should agree that a malevolent choice architect, aware of the power of nudges, could produce a great deal of harm. But the most serious harms tend to come from mandates and bans – from coercion – and not from nudges, which maintain freedom of choice.

11. 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.<sup>26</sup> People might be especially averse to intentional nudges. If we are especially fearful of official mistakes – coming from incompetence or bad motivations – we will want to minimize the occasions for nudging. And if we believe that invisible hand mechanisms promote welfare or freedom, we will not want to disturb their products, even if those products include nudges. But a degree of official nudging cannot be avoided.
12. In this essay, I will offer seven principal conclusions. (a) It is pointless to object to choice architecture or nudging as such. The private sector inevitably nudges, as does the government. We can object to particular nudges, and particular goals of choice architects, and particular forms of choice architecture but not to nudging and choice architecture in general. For human beings (or for that matter dogs and cats and mice), choice architecture cannot be avoided. It is tempting to defend nudging on the part of government by saying that the private sector already nudges (sometimes selfishly, even in competitive markets<sup>27</sup>). On certain assumptions, this defense might be right, but is not necessary, because government is nudging even if it does not want to do so. (b) In this context, ethical abstractions (about, for example, autonomy, dignity, and manipulation<sup>28</sup>) can create serious confusion. We need to bring those abstractions into contact with concrete practices. Nudging takes many diverse forms, and the force of an ethical objection depends on

---

<sup>25</sup> See Ullmann-Margalit, *supra* note.

<sup>26</sup> See Edward Glaeser, *Paternalism and Psychology*, 73 U. Chi. L. Rev. 133, 136-39 (2006).

<sup>27</sup> See Bar-Gill, *supra* note.

<sup>28</sup> A valuable discussion is T. M. Wilkinson, *Nudging and Manipulation*, 61 *Political Studies* 341 (2013). See also Sarah Conly, *Against Autonomy* 30 (2012): “Libertarian Paternalism is manipulative. That is, it does not suggest that we engage in free and open discussion in order to rationally persuade you to change your ways. . . . The point of the nudge is to push you in ways that bypass your reasoning. That is, they use your cognitive biases, like your tendency to go with the default option, to bring about good effects. There is a sense in which they fail to respect people’s decision-making ability.” As explained below, the point of the nudge is not “to push you in ways that bypass your reasoning,” but the concern must be engaged in the context of (some) nudging.

specific form.<sup>29</sup> (c) If welfare is our guide, much nudging is actually required on ethical grounds. (d) If autonomy is our guide, much nudging is also required on ethical grounds. (e) Choice architecture should not, and need not, compromise either dignity or self-government, though imaginable forms could do both. (f) Many nudges are objectionable because the choice architect has illicit ends. If the ends are legitimate, and if nudges are fully transparent and subject to public scrutiny, a convincing ethical objection is far less likely to be available. (g) There is, however, room for such an objection in the case of highly manipulative interventions, certainly if people have not consented to them. The concept of manipulation deserves careful attention, especially because of its relationship to the ideas of autonomy and dignity.

## II. The Diversity of Nudges and the Dangers of Abstraction

13. For purposes of orientation, it will be useful to give a sense of potential nudges that might alter choice architecture.<sup>30</sup> The most obvious consist of default rules, which establish what happens if people do nothing at all.<sup>31</sup> Others include simplification (for example, of applications for job training or financial aid); disclosure of factual information (for example, calorie labels, designed to promote healthier choices, or of the use of “conflict minerals,” understood as minerals used to finance conflicts in the Democratic Republic of the Congo); warnings, graphic or otherwise (for example, on cigarette packages); reminders (for example, of bills that are about to become due or of the need to take medicine at a certain time); increases in ease and convenience (for example, of healthy goods); 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?”); and precommitment strategies<sup>32</sup> (through which people agree, in advance, to a particular course of conduct, such as a smoking cessation program).
14. 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.<sup>33</sup> System 2 can and does err, but System 1 is distinctly

---

<sup>29</sup> Id.

<sup>30</sup> See Cass R. Sunstein, *Nudging: A Very Short Guide*, J Consumer Policy (2014), available in preliminary form at

[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2499658&download=yes](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2499658&download=yes)

<sup>31</sup> Eric Johnson et al., *Defaults, Framing and Privacy: Why Opting In-Opting Out*, 13 Marketing Letters 5 (2002)

<sup>32</sup> See Ian Ayres, *Carrots and Sticks* (2011).

<sup>33</sup> See Daniel Kahneman, *Thinking, Fast and Slow* (2011).

associated with identifiable behavioral biases.<sup>34</sup> Some nudges attempt to strengthen the hand of System 2 by improving the role of deliberation and people's considered judgments – as, for example, through disclosure strategies and the use of precommitment. Other nudges are designed to appeal to, or to activate, System 1 – as in the cases of graphic health warnings.<sup>35</sup> Some nudges work because of the operation of System 1 – as, for example, where default rules have large effects because of the power of inertia.<sup>36</sup>

15. A nudge might be justified on the ground that it helps counteract a behavioral bias,<sup>37</sup> but (and this is an important point) such a bias is *not* a necessary justification for a nudge.<sup>38</sup> Disclosure of information can be helpful even in the absence of any bias. A GPS is useful even for people who do not suffer from present bias, probability neglect, or unrealistic optimism.<sup>39</sup> A default rule simplifies life and might therefore be desirable whether or not a behavioral bias is involved.
16. 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 desirable as a way of promoting simple navigation. To date, there has been far too little attention to the close relationship between navigability and (good) nudges.<sup>40</sup> Insofar as the goal is to promote navigability, the ethical objections are greatly weakened.

---

<sup>34</sup> See *id.*

<sup>35</sup> For a discussion of first amendment issues, see *R.J. Reynolds Tobacco Co. v. U.S. Food & Drug Admin.*, 823 F. Supp. 2d 36 (D.D.C. 2011), *rev'd*, 696 F.3d 1205 (D.C. Cir. 2012).

[http://www.gpo.gov/fdsys/pkg/USCOURTS-dcd-1\\_11-cv-01482/pdf/USCOURTS-dcd-1\\_11-cv-01482-0.pdf](http://www.gpo.gov/fdsys/pkg/USCOURTS-dcd-1_11-cv-01482/pdf/USCOURTS-dcd-1_11-cv-01482-0.pdf)

<sup>36</sup> See Johnson et al., *supra* note.

<sup>37</sup> See Christine Jolls and Cass R. Sunstein, *Debiasing Through Law*, 35 *J Legal Stud* 199 (2006).

<sup>38</sup> This conclusion means that the highly illuminating discussion in RICCARDO REBONATO, *TAKING LIBERTIES* (2012), offers a mistaken definition.

<sup>39</sup> Conly, *supra* note, writes that for those who endorse nudging, the “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.” *Id.* at 30. This is not the assumption that lies behind nudging, though perhaps some nudges can be understood in this way. See below.

<sup>40</sup> Donald Norman, *The Design of Everyday Things* (2002), was in fact a principal inspiration for the discussion of choice architecture in *Nudge*, *supra* note.

17. Nudges can have a substantial effect on both individual lives and social welfare. In Denmark, automatic enrollment in retirement plans has had a much larger effect than substantial tax incentives.<sup>41</sup> In the context of credit cards, disclosure requirements have saved American consumers many millions of dollars.<sup>42</sup> 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 as a significant spike in the short-term cost of electricity.<sup>43</sup> 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.<sup>44</sup>
18. It must be acknowledged that choice architecture can be altered, and 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.<sup>45</sup> With these points, there is no objection to nudges as such; the objection is to the grounds for the particular nudges.
19. For example, an imaginable default rule might skew the democratic process by saying that voters are presumed to vote to support 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 about the supposedly nefarious plans of members of a minority group. Social norms might be used to encourage people to buy unhealthy products. In extreme cases, private or public institutions might try to nudge people toward violence.
20. It must also be acknowledged that the best choice architecture often calls for active choosing.<sup>46</sup> 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,

---

<sup>41</sup> 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), available at <http://www.nber.org/papers/w18565>

<sup>42</sup> Sumit Agarwal et al., *Regulating Consumer Financial Products: Evidence from Credit Cards* (2013), available at <http://www.nber.org/papers/w19484>

<sup>43</sup> See Hunt Alcott, *Social Norms and Energy Conservation*, 85 *J Public Econ* 1082 (2011).

<sup>44</sup> See Eric Bettinger et al., *The Role of Simplification and Information in College Decisions* (2009), available at <http://www.nber.org/papers/w15361>

<sup>45</sup> See Glaeser, *supra* note; Rebonato, *supra* note.

<sup>46</sup> See Cass R. Sunstein, *Deciding By Default*, 162 *U Pa L Rev* 1 (2013).

without imposing any requirement that they do so.<sup>47</sup> 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.

21. Choice architecture should be transparent and subject to public scrutiny, certainly if public officials are responsible for it. In general, regulations should be subject to a period of public comment. If officials alter a default rule so as to promote clean energy or conservation, they should not hide what they are doing. Self-government itself requires public scrutiny of nudges – a form of choice architecture for choice architects. 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, and of nudges that threaten autonomy or dignity. Nations should also treat their citizens with respect, and public scrutiny shows a measure of respect.
22. 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.

### III. “As Judged By Themselves”

23. There have been recurrent ethical concerns about nudging, taken as such.<sup>48</sup> The principal concerns involve autonomy, dignity<sup>49</sup> (sometimes

---

<sup>47</sup> Kessler and Roth, *supra* note.

<sup>48</sup> Edward L. Glaeser, *Paternalism and Psychology*, 73 U. Chi. L. Rev. 133, 136-39 (2006) (offering examples of how individuals’ beliefs and opinions can be manipulated); 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).

<sup>49</sup> See Jeremy Waldron, *It’s All For Your Own Good*, New York Review of Books (2014), available at <http://www.nybooks.com/articles/archives/2014/oct/09/cass-sunstein-its-all-your-own-good/> Consider in particular this question in *id.*: “Deeper even than this is a prickly concern about dignity. What becomes of the self-respect we invest in our own willed actions, flawed and misguided though they often are, when so many of our choices are manipulated to promote what someone else sees (perhaps rightly) as our best interest?”

described as “respect for persons”), manipulation,<sup>50</sup> and learning.<sup>51</sup> In their most ambitious forms, the concerns lead to a conclusion that nudges can or do violate individual rights. Obviously this might be true if their goal is illicit, but some people believe that some kind of violation (for example, an insult to autonomy) can occur even without an illicit goal.<sup>52</sup>

24. It is important to emphasize that the objective of nudging is to “influence choices in a way that will make choosers better off, *as judged by themselves*”<sup>53</sup> (italics in original). In most cases, this standard is simple 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. A purely factual disclosure, or a warning about risks associated with a product, should be designed in light of this standard, though admittedly some disclosure policies are not effective.<sup>54</sup>
25. It must be acknowledged that in some cases, the “as-judged-by-themselves” standard will leave serious ambiguity. Choice architects might not have sufficient information to know whether choosers deem themselves to be better off. It might not be simple to compare (from choosers’ point of view) the various outcomes that stem from different nudges. Nonetheless, the idea of subjective well-being serves as a general lodestar.<sup>55</sup> Certainly choice architects should be focused on the welfare of choosers, rather than their own.<sup>56</sup> (In a well-functioning market system, that focus is essentially guaranteed, at least under highly optimistic assumptions.<sup>57</sup>)
26. There are also hard questions about how to handle the “as judged by themselves” criterion in the face of self-control problems. Suppose that

---

<sup>50</sup> See White, *supra* note.

<sup>51</sup> See Rebonato, *supra* note.

<sup>52</sup> See *id.*

<sup>53</sup> Thaler and Sunstein, *supra* note, at 5.

<sup>54</sup> See George Loewenstein et al., Disclosure: Psychology Changes Everything, 6 Annual Review of Economics 391 (2014).

<sup>55</sup> Hard questions might be raised where subjective well-being departs from objective well-being; I am bracketing those questions. See Martha Nussbaum, *Creating Capabilities* (2011); Amartya Sen, *Development As Freedom* (2000). 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.

<sup>56</sup> I am bracketing cases in which there are third-party effects, which would require an aggregate welfare judgment, not merely a focus on choosers.

<sup>57</sup> Optimistic, not realistic. See Oren Bar-Gill, *Seduction By Contract* (2011).

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 the alcoholic, who wants that beer, or the ex-alcoholic, who is grateful to have been nudged away from alcoholism<sup>58</sup>? It is reasonable to favor the latter, in part on the ground that no ex-alcoholic regrets the “ex”; in such circumstances, the “as judged by themselves” criterion is best understood to refer to the judgment of the person who is no longer in the grip of an addiction. Nonetheless, there can be a thin line between a self-control problem and a legitimate focus on short-term pleasure; the question deserves extended treatment.

27. Recall at this point that choice architecture is inevitable. Any website nudges; so does a cell phone or a computer; so do lawyers and doctors. A doctor can try to present options in a neutral way so as to respect patient autonomy; but that is a form of choice architecture, not an alternative to it. Whenever government has websites, offices, or programs, it creates choice architecture, and it will nudge.
28. Decades of work in behavioral science have specified how human beings depart from full rationality.<sup>59</sup> It would not be correct to say that people are “irrational.” It is more helpful to draw attention to “bounded rationality.” Most obviously, people often lack important information. They are also subject to specific biases. For example, most people tend to be unrealistically optimistic.<sup>60</sup> People also show “present bias,” focusing on the short-term and downplaying the future.<sup>61</sup> People do not deal well with probability, in part because they use heuristics, or mental shortcuts, that sometimes lead them in unfortunate directions.<sup>62</sup>
29. It is true, of course, that in the face of error, education might be the best response. Some people argue in favor of educational interventions in lieu of nudges.<sup>63</sup> In a way, the opposition is confusing; at least some such interventions fit the definition of a nudge, and they are certainly a form of choice architecture. When education is favored, a natural question is this: Favored over what?
30. In some cases, a default rule would be preferable to education, because it would preserve desirable outcomes (again, from the standpoint of choosers themselves) without requiring people to take the functional

---

<sup>58</sup> On some foundational questions, see Jon Elster, *Ulysses and the Sirens* (1984).

<sup>59</sup> See Daniel Kahneman, *Thinking Fast and Slow* (2011).

<sup>60</sup> See Tali Sharot, *The Optimistic Bias* (2011).

<sup>61</sup> For an overview with references, see Sunstein, *Why Nudge?*, *supra* note.

<sup>62</sup> See Kahneman, *supra* note.

<sup>63</sup> See Gerd Gigerenzer, *Risk Savvy* (2014).

equivalent of a course in (say) statistics or finance.<sup>64</sup> For those who purchase cell phones, tablets, and computers, it would be impossibly demanding to insist on the kind of education that would allow active choices about all relevant features. Much of life is feasible because products and activities come with default rules, and people are not required to undergo some kind of instruction before selecting them. There is a recurring question whether in particular circumstances, the costs of education justify the benefits. Default rules may well be best.

31. Though choice architecture and nudging are inevitable, some particular nudges are certainly avoidable.<sup>65</sup> A government might decide not to embark on a campaign to discourage smoking or unhealthy eating. It might refrain from nudging people toward certain investment behavior. To that extent, it is reasonable to wonder whether government should minimize nudging.<sup>66</sup> If we distrust the motives of public officials, or if believe that their judgments are likely to go wrong, we will favor such minimization.<sup>67</sup> Some of the ethical objections to choice architecture and to nudging are best understood as a plea for minimization, not elimination. In fact the plea might well be more precise than that. At least in some cases, it is a claim that government should avoid particular interventions that are taken to be manipulative and hence to compromise both autonomy and dignity.

#### IV. Seven Objections

32. Here is a catalogue of potential objections to nudges as such. (a) Nudges are paternalistic. (b) Some nudges intrude on people's autonomy. (c) Some nudges might be seen as coercive, even if they preserve freedom of choice as a technical matter. (d) Some nudges insult people's dignity; they might be infantilizing; they might treat people as children.<sup>68</sup> The idea of the "nanny state" captures this objection. (e) Some nudges could count as forms of manipulation.<sup>69</sup> It is relevant in this regard that nudging is not always transparent. Consider,

---

<sup>64</sup> See Lauren Willis, *The Financial Education Fallacy*, 101 *Am Econ Rev* 429 (2011).

<sup>65</sup> See Gleaser, *supra* note.

<sup>66</sup> See *id.*

<sup>67</sup> *Id.*

<sup>68</sup> See Joel Feinberg, *Legal Paternalism*, in *Paternalism* 3 (R. Darforius ed. 1983): "If adults are treated as children they will come in time to be like children. Deprived of the right to choose for themselves, they will soon lose the power of rational judgment and decision." This is a claim about deprivation of the right to choose, but it could be adapted to apply to default rules as well.

<sup>69</sup> As we shall see, manipulation is best understood as troublesome, on ethical grounds, if it runs into a foundational commitment of some kind – for example, to autonomy or dignity. I explore it separately because it raises distinctive considerations.

for example, “negative option marketing,” by which people who purchase certain products find themselves enrolled in programs for which they pay a monthly fee.<sup>70</sup> Nudges of this kind might be said to operate “behind people’s backs.” (The concept of manipulation is not self-defining and deserves considerable attention.<sup>71</sup>) (f) Some nudges impede or at least do not promote learning. (g) Choice architects may err, especially when they work for government, and for that reason, it is best to avoid nudging (to the extent that this is possible).

33. It is important not to take these concerns as all-purpose objections to efforts to improve choice architecture. Does any of these objections make sense as applied to initiatives to promote active choosing? To inform consumers of the caloric content of food, to remind people that a bill is due, or to ask people whether they want to enroll in a retirement plan? We might be skeptical about the force of these concerns as applied to the overwhelming majority of real-world nudges. But let us take the objections in sequence.

#### **A. Paternalism**

34. If paternalism is objectionable, it is because it runs afoul of some kind of foundational commitment or principle. For example, it might undermine autonomy.<sup>72</sup> While the term is often used as a freestanding objection, the real complaint is that paternalism, in general or in particular circumstances, violates a principle that people rightly value.
35. 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).<sup>73</sup> A GPS can be so understood, and the same is true for a reminder, a warning, a use of social norms, and a default rule.
36. This is a distinctive form of paternalism in the sense that it is (a) soft and (b) means-oriented.<sup>74</sup> It is soft insofar as it avoids coercion or material incentives, and thus fully maintains freedom of choice.<sup>75</sup> It is

---

<sup>70</sup> See 16 C.F.R. § 425.1 (2012) (regulating the use of prenotification negative-option plans); FTC, Negative Options 2 (2009), available at <http://www.ftc.gov/os/2009/02/P064202negativeoptionreport.pdf> (describing four types of plans that could be classified as negative-option marketing).

<sup>71</sup> See note *supra*.

<sup>72</sup> For different perspectives, see Rebonato, *supra* note; Conly, *supra* note.

<sup>73</sup> Thaler and Sunstein, *supra* note.

<sup>74</sup> See Cass R. Sunstein, *Why Nudge?* (2014).

<sup>75</sup> On some of the complexities here, see *id.*

means-oriented insofar as it does not attempt to question or alter people's ends. Like a GPS, it respects those ends. To those who object to paternalism, the most serious concerns arise in the face of coercion (where freedom of choice is blocked) and when social planners, or choice architects, do not respect people's ends.<sup>76</sup> To this extent, nudges aspire to avoid some of the standard ethical objections to paternalism.

37. Nonetheless, some skeptics object to paternalism as such.<sup>77</sup> Many of those objections point to individual welfare, and to the risk that planners or choice architects will compromise it. 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. A coercive GPS would, in some cases, intrude on people's ends.
38. 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).
39. If 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 really 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.<sup>78</sup>
40. But insofar as a GPS is a guiding analogy, and insofar as freedom of choice is fully maintained, it is not easy to see nudges as objectionably paternalistic. At least some nudges are entirely focused on means. Consider cases in which people are *mistaken about facts* (with respect to the characteristics of, say, a consumer product or an investment). If a

---

<sup>76</sup> This is the fundamental concern in John Stuart Mill, *On Liberty* (2d ed. 1863), reprinted in *The Basic Writings of John Stuart Mill: On Liberty, The Subjection of Women, and Utilitarianism* 3, 11-12 (Dale E. Miller ed., 2002).

<sup>77</sup> Rebonato, *supra* note; 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).

<sup>78</sup> See Rebonato, *supra* note.

nudge informs them, then it is respecting their ends. 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.

41. 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 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 by hypothesis maximizing 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. It is hard question whether the relevant ends are those that the chooser has at that time or at a later time, or whether it is best (as I tend to think) to identify the chooser's ends by focusing on some aggregation or index of selves over a lifetime.
42. Let us bracket the most difficult issues and acknowledge that some forms of choice architecture count as paternalistic. Is that a problem? As we have seen, one reason to reject paternalism involves welfare: Perhaps people are the best judges of what will promote their interests, and perhaps outsiders will blunder (as Mill believed). 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.*"<sup>79</sup> A form of paternalism that maintains liberty, and that is focused on means, is less likely to be objectionable on welfare grounds, certainly if we attend to behavioral biases.
43. A possible response is that even means-oriented nudges will be inadequately informed, at least if they come from government. When public officials design warnings, they might make mistakes; when they produce default rules, they might steer people in bad directions; when they inform people, they might tell them something other than what it is useful to know. If we are deeply skeptical of the good faith and competence of public officials, we will want to minimize official

---

<sup>79</sup> Friedrich Hayek, *The Market and Other Orders*, in *The Collected Works of F. A. Hayek* 384 (Bruce Caldwell ed., 2013).

nudges, and we will prefer choice architecture that comes from invisible hands and spontaneous orders.

44. This view cannot be ruled out in the abstract. It depends on certain assumptions about the risk of government error; whether or not it is right, it is not clear that it should be counted as a distinctly *ethical* objection.<sup>80</sup> Nor it is clear that it should count as an objection to efforts to increase navigability, unless we think that those efforts are themselves likely to be misconceived.
45. In the face of ignorance of fact and behavioral biases, some welfarists are actually drawn to coercive paternalism.<sup>81</sup> When paternalism would improve welfare, welfarists should be inclined to support paternalism. For welfarists, there is a good argument that paternalism, hard or soft, should be evaluated on a case-by-case basis – unless there is some systematic, or rule-welfarist, justification for a principle or presumption against paternalism.<sup>82</sup>
46. There may be good reason for such a presumption, rooted in a judgment that choosers are likely to have better information than choice architects. But in some cases, that judgment is incorrect, because choosers lack, and choice architects have, accurate knowledge of relevant facts. There are serious risks in using this point as a reason for coercion, at least when the interests of third parties are not involved. The advantage of nudges is that they reduce those risks, because people are allowed to go their own way.<sup>83</sup>
47. In the face of missing information, information-providing nudges are a natural corrective. But 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, but that conclusion is not obvious; much depends on the costs of decisions and the costs of errors.<sup>84</sup> Welfarists might well be inclined to favor choice-preserving approaches, on the theory that individuals usually well know what best fits their circumstances, but default rules preserve choice, and the fact that they have a paternalistic dimension should not be decisive against them.
48. Another reason to reject paternalism involves autonomy and the idea of respect for persons. Stephen Darwell writes that the “objectionable

---

<sup>80</sup> See below.

<sup>81</sup> See Sarah Conly, *Against Autonomy* (2012).

<sup>82</sup> See Cass R. Sunstein, *Nudges v. Shoves*, 127 Harv L Rev Forum 210 (2014).

<sup>83</sup> See *id.*

<sup>84</sup> See Sunstein, *Choosing Not to Choose* (forthcoming 2015).

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.”<sup>85</sup> This brings us to the next objection.

## **B. Autonomy**

49. Do nudges intrude on autonomy? Recall that nudges are inevitable, so in a sense, the question is confusing. It appears to be premised on a judgment that existing nudges do not compromise autonomy and that a new nudge, proposed or actual, would introduce nudging where it did not exist before. That is a mistake. The real question is whether particular nudges intrude on autonomy.
50. In any case, autonomy requires informed choices, and many nudges are specifically designed to ensure that choices are informed.<sup>86</sup> In the face of a lack of information, a behavioral bias, or some kind of systematic mistake (by the actor’s own reflective lights), it is hardly clear that a nudge, taken as such, infringes on autonomy, rightly understood.<sup>87</sup> And when social contexts are not navigable, a nudge that improves navigability increases autonomy, rather than undermining it. A GPS does not create a problem from the standpoint of autonomy; nor does a user-friendly computer; nor does an effort to simplify a mortgage or a credit card agreement.
51. 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 people have to make choices everywhere, their autonomy is reduced, if only because they cannot devote attention to those activities that seem to them most worthy of their attention.<sup>88</sup> There is a close relationship between time-management and autonomy. To the extent that nudges reduce the difficulty of time-management, they increase an important form of autonomy.

---

<sup>85</sup> See Stephen Darwell, *The Value of Autonomy and the Autonomy of the Will*, 116 *Ethics* 263, 269 (2006).

<sup>86</sup> See George Loewenstein et al., *Disclosure: Psychology Changes Everything*, 6 *Annual Review Economics* 391 (2014).

<sup>87</sup> See Conly, *supra* note, 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?”

<sup>88</sup> See Sendhil Mullainathan and Eldar Shafir, *Scarcity* (2013).

52. It is nonetheless true that on grounds of autonomy (as well as welfare), the best choice architecture often calls for active choosing. Default rules might intrude on autonomy, certainly if they do not track people's likely (informed) choices. It is important to remember that because of the force of inertia, people might not reject harmful defaults.<sup>89</sup> If so, there is an intrusion on their autonomy, because they will end up with outcomes that they did not specifically select and they do not or will not like.
53. Whether the interest in autonomy calls for active choosing, as opposed to reliance on a default rule, depends on the circumstances. Often active choosing is best and should be built into the relevant choice architecture – but along some dimensions, default rules can be 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.<sup>90</sup> 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.<sup>91</sup> Preservation of freedom of choice goes a long way toward respecting autonomy. With respect to autonomy, the real problem lies in the possibility of manipulation; I will turn to that problem below.

### **C. Coercion**

54. If choice architects coerce people, they are no longer merely nudging. But skeptics might again emphasize 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.<sup>92</sup>
55. We should 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 even though people have not affirmatively consented to the actions that led to them. Choice architects need to take account of that risk. But so long as freedom of choice is maintained, and real rather than formal, coercion is not involved.

### **D. Dignity**

---

<sup>89</sup> See Rebonato, *supra* note.

<sup>90</sup> See Cass R. Sunstein, *Choosing Not to Choose*, 64 *Duke LJ* 1 (2014).

<sup>91</sup> See *id.*

<sup>92</sup> For an excellent discussion, see Rebonato, *supra* note.

56. The antonym of coercion is freedom; the antonym of dignity is humiliation.<sup>93</sup> As we shall see, this may be the most interesting objection of all, especially when it is combined with a concern about manipulation.<sup>94</sup>
57. There are large questions about the place of dignity in ethics.<sup>95</sup> On one (less than conventional) 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.
58. A good welfarist should also acknowledge that an offense to dignity is qualitatively distinct; 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 loss, to be weighed against the other goods that are at stake. Suppose, for purposes of argument,<sup>96</sup> that a graphic and highly emotional appeal, triggering strong emotions (System 1) in order to discourage people from smoking, is plausibly seen as an offense of dignity – as a way of treating smokers disrespectfully. (Some smokers might so regard such an appeal and object for that reason.) A welfarist might be willing to support the emotional appeal, notwithstanding the relevant loss, if it saves a significant number of lives.
59. On another 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. It should not be permitted unless (perhaps) it has an overwhelmingly strong justification. If we endorse this view, it is especially important to ask whether nudges offend human dignity.
60. 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

---

<sup>93</sup> See Avishai Margalit, *The Decent Society* (1998).

<sup>94</sup> See Waldron, *supra* note.

<sup>95</sup> See Margalit, *supra* note; Michael Rosen, *Dignity: Its History and Meaning* (2012); Charles Beitz, *Human Dignity in the Theory of Human Rights*, 41 *Phil & Pub. Affairs* 259 (2013); Thomas Christiano, *Two Conceptions of Human Dignity as Persons* (unpublished manuscript 2008).

<sup>96</sup> For discussion, see Kyle Rozema, *Economic Theory Lost in Translation: Will Behavioral Economics Reshape the Compelled Commercial Speech Doctrine?* (2013), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2511604](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2511604)

information is useful and not based on a false and demeaning belief that people need it. Nudges that increase navigability do not offend dignity.

61. True, it might be 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 can be found in adult life as well. If people are informed of the same thing every hour or even every day, they might legitimately feel that their dignity is not being respected. If people are constantly reminded that a due date is coming, they might feel as if they are being treated like children.
62. If choice architects refer to social norms, to let people know what most people do, they are not likely to be humiliating anyone.<sup>97</sup> In some cases, however, repeated invocations of social norms might run into a concern about dignity. Warnings can run into the same concern insofar as they are repetition or condescending, or (are meant to) trigger strong emotions instead of merely giving people a sense of factual realities.<sup>98</sup>
63. Here as well, there is no objection to the relevant nudges in the abstract, but there is an objection to imaginable nudging. At the same time, it must be emphasized that outside of exotic hypotheticals, the relevant offense to dignity – coming from unwelcome and numerous reminders – 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.<sup>99</sup>
64. 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 defaults for the phone and the contractual arrangement, nothing need be amiss in terms of dignity.
65. But we could imagine harder 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. The reasonable response is: Why shouldn’t a free people be asked to choose what they want? Or suppose that a government specified a “default exercise plan” for adults, so that they would be assumed to engage in certain activities unless they expressly opted out. People might offer the same reasonable response, perhaps with considerable agitation. A more modest and more realistic proposal is

---

<sup>97</sup> See Alcott, *supra* note.

<sup>98</sup> See *id.*

<sup>99</sup> See note *supra*.

that doctors should enroll smokers, by default, into some kind of smoking cessation plan (subject to opt-out). Might automatic enrollment offend dignity?

66. Note that default rules of these kinds might be objectionable for both welfarists and nonwelfarists. Welfarists might want to focus on people's subjective feelings. Their belief that they are being treated as children, and their objection to that treatment, would count in the assessment. Nonwelfarists would insist that the offense to dignity is objectionable even if it has some kind of welfarist justification. (There is a question whether and when nonwelfarists would be willing to allow welfarist consideration to override the objection.)
67. 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). But it would be a mistake to use extreme situations, or 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 – or in science-fictional accounts of “nudge world”<sup>100</sup> -- than in the context of the vast majority of real-world cases in which default rules are at work. Admittedly, the objection must be taken seriously in some real-world contexts.

### E. Manipulation

68. To deal with this objection, we need to say something about the complex idea of “manipulation.”<sup>101</sup> An initiative does not count as manipulative merely because it is an effort to alter people's behavior. If you warn a driver that he is about to drive into a ditch or get into a crash, you are not engaged in manipulation. The same is true if you remind someone that a bill is due or that a doctor's appointment is upcoming.
69. It is not clear that the idea of manipulation can be subject to a simple definition, or a statement of necessary and sufficient conditions.<sup>102</sup> It might be an umbrella concept for an assortment of practices. But we might begin by saying that manipulation exists when *someone tries to alter people's behavior in a covert way, by deceiving them about, or*

---

<sup>100</sup> See Waldron, *supra* note.

<sup>101</sup> For helpful discussion, see Wilkinson, *supra* note; White, *supra* note.

<sup>102</sup> An especially good exploration is Wilkinson, *supra* note.

*hiding, or even failing to disclose, a relevant aspect of the interaction.*<sup>103</sup>  
The idea of “relevant aspect” should be understood by reference to what matters to people. If someone has persuaded you to buy a new cell phone by falsely describing an important feature of the product, you have been manipulated. Or if someone has persuaded you to buy a home without disclosing the fact that the heating system does not work, you can claim to have been manipulated.

70. A lie is a defining example of manipulation. Deceptive behavior counts as manipulative as well, even if no one has actually spoken falsely. If you imply that certain food is unhealthy to eat, when it is not, you are engaged in manipulation.
71. An action might generally be counted as manipulative if it *lacks transparency* – if the role or the motivation of the choice architect is hidden or concealed. In the pivotal scene in *The Wizard of Oz*, the wizard says, “Pay no attention to the man behind the curtain.” The man behind the curtain is of course a mere human being who is masquerading as the great Wizard. If choice architects conceal their own role, it seems fair to charge them with being manipulative.
72. An action can also be counted as manipulative if it attempts to influence people subconsciously or unconsciously, in a way that undermines their capacity for conscious choice. Consider the suggestion that “manipulation is intentionally and successfully influencing someone using methods that pervert choice.”<sup>104</sup> Of course the term “pervert choice” is not self-defining; it might well be 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.”<sup>105</sup> The objection applies to lies, which attempt to alter behavior by appealing to falsehoods rather than truth (where falsehoods would enable people to decide for themselves). In harder cases, the challenge is to concretize the ideas of “subverting” and “insulting.”
73. Subliminal advertising may 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

---

<sup>103</sup> See *id.*

<sup>104</sup> See note *supra*.

<sup>105</sup> *Id.* at

involuntary hypnosis does not have much appeal. The question is whether taboo practices can shed light on interventions that can command broader support.

74. On one view, nudges generally or frequently count as manipulative. Sarah Conly suggests 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.”<sup>106</sup>
75. This is a strong charge, and it is not fairly leveled against most kinds of nudges. Recall some examples: disclosure; reminders; warnings; default rules<sup>107</sup>; simplification. Some forms of choice architecture are rooted in an acknowledgement that human beings suffer from bounded rationality, but they do not appeal to “irrationality,” or reflect a judgment that “people can’t generally make good decisions when left to their own devices.”
76. But consider some testing cases, where Conly’s charge 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.<sup>108</sup> (b) Choice architects might be alert to *framing effects* and present information accordingly.<sup>109</sup> (1) They might enlist loss aversion, suggesting that if people *decline* to engage in certain action, they will *lose* money, rather than suggesting that if they *engage* in certain action, they will *gain* money. (2) 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 order options – in a cafeteria or on a form – so as to make it more likely that people will make certain choices.

---

<sup>106</sup> Conly, *supra* note, at 30.

<sup>107</sup> To be sure, some default rules might be selected on the ground that because of a behavioral bias, the chooser would not chose well.

<sup>108</sup> See Christine Jolls, Product Warnings, Debiasing, and Free Speech: The Case of Tobacco Regulation, 169 *J Institutional and Theoretical Economics* 53 (2013).

<sup>109</sup> See *Perspectives on Framing* (Gideon Keren ed. 2010).

77. 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 any effort to subvert or to insult people's decision-making powers? 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 eliminate the charge of manipulation? In cases of this kind, the answer is not self-evident.
78. 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.<sup>110</sup> But what if the warning is explained, in public, on exactly that ground? What if a warning is introduced and justified as effective, because it appeals to people's emotions, and thus saves lives? What if it is welcomed by the relevant population – say, smokers – for exactly that reason? Similar questions might be asked about strategic uses of framing effects, social norms, and order effects. T. M. Wilkinson contends, plausibly, 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.”<sup>111</sup>
79. If government is targeting System 1 – perhaps through framing, perhaps through 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 a range of past and current manipulations – including advertising and social norms – have influenced people to become smokers.
80. If this 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, and appeal to their deliberative capacities. There is no obvious answer to the resulting debates. Those who are committed to welfarism might have a different view from those who are committed to some form of deontology.
81. It is not implausible to say that even with full transparency, at least some degree of manipulation is involved whenever a choice architect is

---

<sup>110</sup> See note supra. 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? See note supra.

<sup>111</sup> Wilkinson, supra note, at 345.

targeting emotions or seeking a formulation that will be effective because of its interaction with people's intuitive or automatic thinking (System 1). But there are degrees of manipulation, and there is a big difference between a lie and an effort to frame an alternative in an appealing light.

82. In ordinary life, we would not be likely to accuse our friends or loved ones of manipulation if they offered a smile or a frown if we said that we were seriously considering a particular course of action. It would be an expansive understanding of the word "manipulation" if we used it to cover people who characterized one approach as favored by most members of our peer group, or who emphasized the losses that might accompany an alternative that they abhor. Actions that are conceivably 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.
83. We have seen most nudges are not manipulative in any relevant sense. But to the extent that some of them can be counted as such, the force of the objection or concern depends on the degree of any manipulation. We might well insist on an absolute or near-absolute taboo on lying or deception on government's part, for welfarist or nonwelfarist reasons. But surely we should be more lenient toward emotional appeals and framing. One question is whether such approaches produce significant welfare gains. If a graphic health warning saves many lives, it is unacceptable if and because it can be counted as a (mild) form of manipulation? A welfarist would want to make an all-things-considered judgment about the welfare consequences.
84. It is true that some people, focused on autonomy as an independent good, would erect a strong and perhaps conclusive presumption against defining cases of manipulation.<sup>112</sup> But I hope that I have said enough to show that the modest forms discussed here strain the boundaries of the concept -- and that it would be odd to rule them off-limits.

## F. Learning

85. Choice-making is a muscle, and the ability to choose can be strengthened through exercise. If nudges would make the muscle atrophy, we would have an argument against them. Here too, it is necessary to investigate the particulars.

---

<sup>112</sup> Cf. White, *supra* note.

86. Active choosing and prompted choice hardly impede learning. Nor do information and reminders. On the contrary, they promote learning. Here the point is plain and the evidence is compelling: Nudges of this kind exercise the choice-making muscle, rather than the opposite.<sup>113</sup>
87. With respect to learning, the real problem comes from default rules. It is possible to say that active choosing is far better than defaults, simply because choosing promotes learning.<sup>114</sup> 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.<sup>115</sup> If people are defaulted into certain outcomes, they do not add to their stock of knowledge, and that may be a significant lost opportunity.
88. The argument for learning depends on the setting. For most people, it is not important to become experts in the numerous decisions that lead to default settings on computers and 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. 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.

### **G. Biased Officials**

89. Choice architects are emphatically human as well, and potentially subject to behavioral biases; to say the least, they are often unreliable. 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.<sup>116</sup> 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 be subject to important biases – suffering, for example, from 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.

---

<sup>113</sup> See Benjamin York and Susanna Loeb, *One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers* (2014), available at <http://www.nber.org/papers/w20659>

<sup>114</sup> Cass R. Sunstein, *Choosing Not To Choose* (forthcoming 2015).

<sup>115</sup> See *id.*

<sup>116</sup> Rebonato, *supra* note.

90. 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.<sup>117</sup> No one should deny that proposition, which argues in favor of choice-preserving approaches. If choice architects blunder, at least it can be said that people can go their own way.
91. 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 generally best to rely on free markets and to trust in invisible hand mechanisms. If so, they would select (or accept) choice architecture that reflects those mechanisms.
92. This idea raises many conceptual and empirical questions, which I will not engage here. The question is whether it is so abstract, and so rooted in dogmas, that it ought not to command support. To be sure, free markets have many virtues. But in some cases, disclosure, warnings, and reminders can do far more good than harm.<sup>118</sup> 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 both welfare and freedom.

## **VI. What Do People Think?**

93. I conclude with three empirical findings. These findings cannot dispose of the ethical questions; findings about people's assessments of such questions are not decisive. But they do shed some light on the complexity of those assessments.

### **A. Transparency About Nudging**

94. An obvious question: If people are explicitly informed that they are being nudged, does their behavior change? An important study finds

---

<sup>117</sup> See Cass R. Sunstein, *Nudges vs. Shoves*, 127 Harv. L. Rev. Forum 210 (2014).

<sup>118</sup> See Oren Bar-Gill, *Seduction by Contract* (2012).

that it does not, at least in the important context of end-of-life care.<sup>119</sup> In that context, as elsewhere, the default rule matters: A “no heroic measures” default produces different outcomes from a default that calls for such measures.<sup>120</sup> By itself, that finding is not surprising. For present purposes, the more significant finding is this: *When people are specifically informed that a particular default rule has been put in place, that information has essentially no effect on what people do.*<sup>121</sup> At least in the context of end-of-life care, the effect of a default is not weakened when people are told that a default was chosen because it is usually effective.

95. We do not yet know if the same finding would be made in other contexts or with somewhat different “nudge warnings.” Suppose that people were told that they were being defaulted into being organ donors because default rules tend to stick. It is imaginable that some people would rebel – and opt out. Indeed, the framing of the warning (“you are about to be nudged”) might well matter, with more provocative framing producing increases in “reactance.”<sup>122</sup> Nonetheless, the context of end-

---

<sup>119</sup> See George Loewenstein, Cindy Bryce, David Hagmann & Sachin Rajpal, Warning: You Are About To Be Nudged 17 (Mar. 29, 2014) (unpublished working paper) (<http://ssrn.com/abstract=2417383>) (“[I]nforming people about default interventions in advance . . . does not significantly diminish the impact of defaults on expressed preferences in advance directives.”).

<sup>120</sup> Id.

<sup>121</sup> Here is the instruction, id.:

The specific focus of this research is on ‘defaults’ -- decisions that go into effect if people don't take actions to do something different. Participants in this research project have been divided into two experimental groups.

If you have been assigned to one group, the Advance Directive you complete will have answers to questions checked that will direct health care providers to help relieve pain and suffering even it means not living as long. If you want to choose different options, you will be asked to check off different option and place your initials beside the different option you select.

If you have been assigned to the other group, the Advance Directive you complete will have answers to questions checked that will direct health care providers to prolong your life as much as possible, even if it means you may experience greater pain and suffering.

<sup>122</sup> See Sharon Brehm and Jack Brehm, Psychological Reactance: A Theory of Freedom and Control (1981); Louisa Pavey & Paul Sparks, *Reactance, Autonomy and Paths to Persuasion: Examining Perceptions of Threats to Freedom and Informational Value*, 33 MOTIVATION & EMOTION 277 (2009).

of-life care is not exactly uncontroversial, and it is noteworthy that even in that context, a disclosure of the chosen default, and the fact that it might have been otherwise, did not alter behavior.

96. A warning, a prompted choice, or a reminder is not exactly mysterious; people are certainly aware that the goal is to nudge. We can imagine situations in which warnings and reminders themselves backfire; perhaps people will do exactly what they are warned or reminded not to do. But in the absence of unusual circumstances, there is no reason to think that warnings and reminders will not have their intended effect.<sup>123</sup>

### **B. Politics Matters**

97. Political values greatly matter to people's assessment of nudges.<sup>124</sup> Casual observation suggests that many people seem to like or dislike nudges because of their perceived merits, not because they are nudges. Recent evidence suggests that the observation is correct.<sup>125</sup>
98. In short, people are more likely to raise ethical objections not because they object to nudges as such, but because the particular nudge runs up against their substantive policy preferences. The result is "partisan nudge bias," as "people find nudges more ethically problematic when they are applied to policy objectives they oppose, or when applied by policymakers they oppose, while they find the same nudges more acceptable when they are applied to political objectives they support or by policymakers they support."<sup>126</sup>
99. It follows that people who are right-of-center are less likely to approve of nudges that seem to have a distinctive left-of-interest motivation; they are more likely approve of nudges if they are right-of-center. This finding does not mean that certain nudges cannot be objectionable as such (and regardless of the direction in which they nudge). But it does suggest that the perceived strength of an ethical objection, and even a perception that such an objection exists, may well depend on the substance.

### **C. Nudging System 1**

---

<sup>123</sup> On reminders, see note supra.

<sup>124</sup> David Tannenbaum et al., *On the Misplaced Politics of Behavioral Policy Interventions* (unpublished manuscript 2014).

<sup>125</sup> *Id.*

<sup>126</sup> *Id.*

100. Consider an intriguing finding: People care whether nudges appeal to System 1 or System 2 – but not always, and not always a great deal.<sup>127</sup> In other words, people are more comfortable with nudges that target deliberative capacities than more automatic or unconscious ones, but the difference is not dramatic, and in some circumstances, people do not object to “System 1 nudges.” Notice, for example, the difference between two scenarios, designed to obtain people’s judgments about possible approaches to retirement savings.<sup>128</sup>

*(a) The new design works like this—with every annual salary increase you are provided information in the form of a series of icons representing tropical beaches that shows how much extra leisure you are likely to be able to afford during your retirement by investing different percentages of your increased salary; larger investments now translate into more retirement savings later. You can still choose to keep the entire salary increase instead of investing it, but the information provided results in a subconsciously-driven bias towards investment; in other words, the decision to invest is made more likely as a result of subconscious deliberation. Studies have shown that implementing this policy leads to an increase in retirement savings.*

*(b) The new design works like this—with every annual salary increase you are provided information in the form of a detailed table of your earnings that shows how much extra money you are likely to have during your retirement by investing different percentages of your increased salary; larger investments now translate into more retirement savings later. You can still choose to keep the entire salary increase instead of investing it, but the information provided results in a consciously-driven bias towards long-term investment; in other words, the decision to invest is made more likely as a result of conscious deliberation. Studies have shown that implementing this policy leads to an increase in retirement savings.*

101. Such questions allow for a test of this hypothesis: *people object to nudges that appeal to unconscious or subconscious processes.* There is evidence that people do indeed show a general aversion to nudges that have that characteristic.<sup>129</sup> One reason is that such appear to be less manipulative, because they engage higher-order thinking. Another

---

<sup>127</sup> For relevant discussion, see Gidon Felsen et al., Decisional Enhancement and Autonomy: Public Attitudes Toward Overt and Covert Nudges, 8 Judgment and Decision Making 203 (2012).

<sup>128</sup> Id.

<sup>129</sup> Id.

reason may be that when nudges appeal to unconscious or subconscious processes, they might seem to offend individual dignity. But the difference in reactions, while real, varies across contexts and in general, is fairly described as moderate.<sup>130</sup>

102. Note, moreover, the suggestion that “covertly influencing decision processes such that the resulting decision is *aligned with higher-order* desires may actually enhance autonomy, especially in situations in which the target population is known to want help with a given behavior.”<sup>131</sup> With this point in mind, we might speculate that people would be more receptive to nudges that affect unconscious processes, or System 1, when the relevant group wants and needs help. And indeed, some evidence suggests that when people believe that some kind of behavioral bias – such as a self-control problem -- is genuinely responsible for welfare losses, they become more receptive to nudges that target unconscious or subconscious processes.<sup>132</sup> When people liked (and believed they needed) the nudge on the merits, they are more likely to favor it even if it targets unconscious processes, at least in some contexts.<sup>133</sup>

## V. Conclusion

103. Nudges and choice architecture cannot be avoided, and so it is pointless to object to them or to attempt to wish them away. But intentional changes in choice architecture, deliberately made by choice architects, can indeed run into ethical concerns – most obviously, where the underlying goals are illicit. Indeed, a concern about illicit goals underlies many of the most plausible objections to (some) nudges.

---

<sup>130</sup> Id. In one setting – involving productivity – people showed no preference at all for nudges that affect conscious processing. The “System 1 nudge” took this form, *id.* at 213:

The software works like this—whenever you are working on the project, the program slows down your Internet browsing speed for non-work related websites. You can still choose to view whatever content you would like, but the inconvenience of slower browsing results in a subconsciously-driven bias towards work-related content; in other words, the decision to spend time productively is made more likely without the need for conscious deliberation. Studies have shown that using this software leads to higher rates of productivity, and therefore higher pay as per the contract.

<sup>131</sup> Id. at 207. See in particular the discussion of the healthy eating scenario in *id.* at 206.

<sup>132</sup> Id.

<sup>133</sup> Id. at 206, again with reference to the healthy eating scenario.

104. Where the goals are legitimate, an evaluation of ethical concerns needs to be made with close reference to the context. Disclosure of accurate information, reminders, and (factual) warnings are generally unobjectionable. If nothing is hidden or covert, nudges are less likely to run afoul of ethical constraints, not least because and when they promote informed choices.
105. Default rules frequently make life more manageable, and it does not make much sense to reject such rules as such. At the same time, it must be acknowledged that active choosing might turn out to be preferable to default rules, at least where learning is important and where one size does not fit all.
106. It is also true that some imaginable nudges can be counted as forms of manipulation, raising objections from the standpoint of both autonomy and dignity. That is a point against them. But the idea of manipulation contains a core and a periphery, and some interventions go beyond the periphery. Even when nudges target System 1, it might well strain the concept of manipulation to categorize them as such (consider a graphic warning, or a use of loss aversion in an educational message). If they are fully transparent and effective, if their rationale is not hidden, and if they do not limit freedom of choice, they should not be ruled out of bounds on ethical grounds, whatever the foundations of our ethical commitments.