Choice Architecture: A Mechanism for Improving Decisions While Preserving Liberty?

J.S. Blumenthal-Barby, PhD

I. INTRODUCTION

Extensive research in the social sciences has shed light on two important phenomena. One is that human decision-making is full of predicable errors and biases that often lead individuals to make choices that defeat their own ends. For example, individuals incorrectly predict what will make them happy in the future and are biased towards short-term consequences over long-term ones despite attaching more importance to their long-term goals. Let us call this the "bad choice phenomenon." The second phenomenon is that individuals' decisions and behaviors are powerfully shaped by their environment in logically irrelevant ways; they are impacted by the order in which options are presented, or tone in which they are presented, and even the smell in the room. Let us call this the "influence phenomenon."

Some have argued that it is ethically defensible that the influence phenomenon be utilized to address the bad choice phenomenon.[1] They propose that "choice architects" learn about the various ways to influence choices and then work to design environments, broadly construed, that

---

[1] R. Thaler and C. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New Haven and London: Yale University Press, 2008). This argument, spearheaded by Thaler and Sunstein's has become extremely popular in the US and the UK. Often labeled "libertarian paternalism," the program is being integrated into policymaking and millions of dollars are being dedicated to investigations of how these findings can improve people's health, financial, and environmental choices. Despite its integration into policy, a full-throated defense of the program and analysis of its limitations and boundaries has not been given, which is the aim of this paper. I would also like to comment on the terminology used in this paper. I avoid using the term "libertarian paternalism" in this paper for two reasons. First, I think the term was a bad choice because "paternalism" is *by definition* interfering with someone's liberty [to make them better off], and Thaler and Sunstein want to claim that their efforts to make people better off are *not* interferences with liberty. Second, avoiding provocative or misleading terminology such as "libertarian paternalism" or "nudging" (another term that has been used in this context) allows me to analyze the issues at stake directly without engaging in semantic debates. I do use "choice architecture," as a descriptive term, which I define in the Introduction section of this paper.

promote choices that make people better off. Specifically, choice architects can influence individuals by arranging the order or tone of the presented options, offering incentives, utilizing social norms, offering vivid examples or representations, creating an anchor for thought processes, having them write down an intention or action plan, engaging them in commitment strategies, creating a default, and mapping choice consequences to something that they care about such as money.

Ethical arguments for creating choice environments that lead people to make better choices revolve around two claims: (1) it makes people better off, and (2) it does so in a way that is entirely compatible with individual liberty. The goal of my paper is to examine these two claims and to see if they hold water. I will first turn to the conceptual and normative concerns with the claim that choice architecture makes people "better off," and then in the second half of the paper I will turn to the soundness of the claim that choice architecture is compatible with liberty.

II. MAKING PEOPLE BETTER OFF

The claim that choice architecture can serve as a mechanism for making people better off carries several ambiguities and objections, which I will explore. I will address the following five issues: (a) what is meant by "better off" and whether choice architects rely on a perfectionist account of well-being, (b) who exactly is the intended recipient(s) of the motivated choice architecture, (c) how to handle the concern that choice architects are subject to biases and errors that affect their ability to influence people beneficently, (d) how to handle the concern that choice architects are subject to motivational problems such as malevolence or indifference that affect their ability to influence people beneficently, and (e) whether choice architects are subject to epistemological deficiencies that affect their ability to influence people beneficently.

A. What Does "Better Off" Mean?

The claim that choice architecture can be used to make people better off raises the question of better off by whose or what standards? Proponents of using choice architecture often claim that the answer is: by an individual's *own* standards. In other words, they reject the promotion of any objective or perfectionist standard of wellbeing, and instead rely on a subjective informed desire account of wellbeing.[2] The issue is complicated, however, when we (1) consider not just what proponents claim to be doing, but what they actually are doing based on examples they give, and (2) realize the difficulties in determining what will make people better off "from their own point of view."

Turning to the first point, Gregory Mitchell notes that even though proponents of choice architecture "…do not choose one particular conception of the good that the planner should try to achieve, most of their illustrative examples involve making people healthier or wealthier."[3] This is in fact true when one canvasses the various areas in which choice architecture strategies have been proposed: organ donation, saving money for retirement, promoting healthy foods, environmental conservation, increasing exercise, increasing medication compliance, smoking cessation, and increasing productivity.[4] Indeed, there is a sense of an underlying perfectionist standard of good, namely health and wealth, to which the various nudges direct the masses. A perfectionist standard of the good holds that what is good for a person is fixed independently of her attitudes towards it; it promotes an ideal way of living that people can achieve to a greater or lesser extent, and allows for comparisons regarding the quantity of the good that people have.[5] But it is important to realize that this sort of perfectionism is not what proponents of choice

---

[2] Thaler and Sunstein, *Nudge*, p. 5
[3] G. Mitchell, "Libertarian Paternalism Is an Oxymoron," *Northwestern University Law Review,* 99 (2005), 21
[4] For a thorough litany of examples see Thaler and Sunstein's *Nudge* and the blog, nudges.org
[5] R. Arneson, "Perfectionism and Politics," *Ethics*, 111 (2000), 38

architecture have in mind. Their concerns lie not with choices in favor of goods other than health and wealth per se, but instead with choices that people would not have made had they paid full attention, possessed, understood, and unbiasedly weighed all of the relevant information, and exhibited self-control.[6] Designing choice environments that promote health and wealth is justified not because health and wealth are goods independently of anyone's attitude toward them, but because (1) the vast majority of people do prioritize them,[7] and (2) health and wealth are instrumental for the achievement of most other goods, regardless of what goods people aim to achieve.

This brings us to the second complication to the choice architect's claim to promote people's *own* conceptions of "better off." It is difficult to conclude what will make people better off, even from their own point of view. As it turns out, individuals are horribly poor assessors and predictors of what will make them better off.[8] Moreover, there is often a tension between short-term and the long-term betterment of self; which "self" do we use for the metric? Advocates of choice architecture have recognized the difficulties inherent in claiming to design choice environments that promote what people themselves view as making them better off. For example, Thaler, Sunstein, and Balz write, "…a little thought reveals that this is difficult …to implement...experiments prove that what [people] choose depends on the order in which the items are displayed. What, then, are the true preferences of the [people]? What does it mean to try to devise a procedure for what the [people] would choose 'on their own'?"[9] The claim that choice architects will direct people towards "their own" standards of "better off" begins to look

---

[6] Thaler and Sunstein, *Nudge,* p. 5
[7] Note the difference between this premise and the premise that health and wealth are things that any rational person prefers more rather than less of (a Rawlsian primary goods premise).
[8] D. Gilbert and T. Wilson, "Miswanting: Some Problems in the Forecasting of Future Affective States," in J. Forgas (ed.), *Feeling and Thinking: The Role of Affect in Social Cognition* (Cambridge University Press, 2001), pp. 178-196
[9] R. Thaler, C. Sunstein, and J. Balz, "Choice Architecture," Available at SSRN: http://ssrn.com/abstract=1583509 (2010), 2

more opaque and imprudent.[10] One ready reply is that we should direct people towards *their own* standards of better off *as determined when* they have paid full attention, possessed, understood, and unbiasedly weighed all of the relevant information, and exhibited self-control. But, of course, the problem with this subjective informed desire approach is that it is an idealized hypothetical that cannot guide us. The solution, I think, is for proponents of choice architecture to admit that choice environments are designed to direct people towards generalizable goods, relying on evidence that certain goods simply do make people better off and happier long term. In other words, while it is true that they are not directing 'individual x' towards what 'individual x' chooses on her own, they are directing 'individual x' towards goods that tend to make people happier or better off as judged by the people who have them.

There is one final issue that I would like to raise with respect to the definition of "better off." At the beginning of the paper I pointed out that choice architects want to use the "influence phenomenon" to address the "bad choice phenomenon" and help people make better choices. One might accuse choice architects of conflating "bad" with "not best" or "less than optimal" and wonder why the goal should be to get people to make the best or most optimal choices. When I referred to "bad choices" at the beginning of the paper, I referred to examples of people defeating their own ends through errors in predicting what will make them happy or through bias towards short term consequences despite the long term ones being more important to them. Yet,

---

[10] J. Halpern and R. Arnold, "Affective Forecasting: An Unrecognized Challenge in Making Serious Health Decisions," *Journal of General Internal Medicine*, 23 (2008), 1708-1712. One might draw a distinction between preferences and underlying values (e.g., I may prefer chocolate cake now but that does not mean that I value it as values are endorsed over time), and argue that while people often poorly come to conclusions about what will make them better off assessing and predicting their preferences, they do a good job about concluding what will make them better off from by assessing and predicting their values. This is not so. People do an equally poor job of assessing and predicting their values (not to mention how poorly they assess which preferences are most in line with their values and behaving accordingly). For example, people don't just mispredict preferring spaghetti for breakfast tomorrow (D. Gilbert, M. Gill, and T. Wilson, "The Future is Now: Temporal Correction in Affective Forecasting," *Organizational Behavior In Human Decision Processes* 88 (2002), 430-444, they mispredict that they will move from valuing independence to valuing time with family.

in the preceding section where I analyzed the objective, outcome based, perfectionist conception of good and bad choices I pointed out that choice architects argue that what is bad about certain choices is not the choice itself but that people choose bad things without having paid full attention, possessed, understood, and unbiasedly weighed all of the relevant information, and exhibited self-control. Defeating one's own ends seems fairly easily categorized as a bad choice, whereas choosing without full understanding, unbiased weighing of all options, and impeccable self-control does not seem properly characterized as a *bad* choice, perhaps less than optimal but not bad.

There is some ambiguity as to whether the objection lies in trying to improve people's choices to be the absolute best, or if the objection is more conceptual or semantic relating to the characterization of choices as bad. My suspicion is that it is more about the conceptual point of conflating "less than optimal" with "bad," and those engaged in the discourse of this debate should be mindful of this point, because I think it is a fair one. I do not, however, see the objection to improving people's choices. Of course, there are objections to improving choices in ways that interfere with liberty or autonomy, which include objections about the ability of choice architects to improve people's choices (I will address these objections later), but this is different from an objection to improving choice as such.

B. Who is Made "Better Off?"

Choice architects aim to design choice environments that make "people" better off, but as Salvat has pointed out, often a single action has multiple opposing consequences on differing populations. He writes, "The director of the cafeteria may, for instance, decide to promote

gluten-free dishes out of sympathy for those who are allergic to it, at the expense of all those who are not. On the other hand, she could also decide to promote iron-supplemented food because it improves most people's diet even though it is highly contraindicated for those who suffer from hemachromatosis."[11] Similarly, most people may be made better off by placing desserts in the back of the cafeteria, but an individual with anorexia is not.

Once the conceptual un-clarity is resolved about the intended recipient(s) of the benefits of choice architecture, then there are normative questions about how to balance the benefits to some groups with the lack of benefits, or even harms, to others. There are some standards that can be employed to guide this balance. One standard is to help those who need it most, while imposing minimal costs on others.[12] For example, a timer on a sun lamp that switches the lamp off after a certain default amount of time would be an instance of choice environment design that helps those who need it most (e.g., those who are extremely forgetful and as a result would burn badly), while imposing minimal costs on those who do not (e.g., the cost of having to turn the sun lamp on again to tan longer). A second instance is requiring a cooling off period before a divorce is granted, which would help those who need it most (e.g., those who lose their temper and do very regretful things), while imposing minimal costs on those who do not (e.g., the cost of having to wait a short period to divorce). Another standard to guide the balance is a Rawlsian standard that the choice environment should be designed to benefit the group that most people would agree should be benefited if they were deciding as free and equal citizens under a veil of ignorance.[13] So, in the case of the choice architect trying to decide whether she should design the cafeteria to nudge people towards nutrient enriched foods or towards iron-free foods, the

---

[11] C. Salvat, "Is Libertarian Paternalism an Oxymoron? A Comment on Sunstein and Thaler," *HAL Working Papers Series* hal-00336528_v1, (2008), 11
[12] Thaler and Sunstein, *Nudge*, p. 242
[13] Rawls, *A Theory of Justice*, (Cambridge, MA: Harvard University Press, 2008)

assumption is that most people would, under a veil of ignorance, agree to position the nudge in a way that makes the majority of people who can benefit from nutrients better off.[14]

C. The Problem of Cognitive and Affective Biases and Errors Affecting the Choice Architect's Judgment

Having dealt with the questions of what is meant by "better off" and who is purportedly made "better off" by the choice architect, I now turn to three additional challenges to the claim that choice architecture actually will make people better off: (1) cognitive and affective biases and errors affecting the choice architect's judgment, (2) motivational problems such as indifference, conflicts of interest, or malevolence affecting the choice architect's judgment, and (3) epistemological deficiencies affecting the choice architect's judgment.

Turning to the first challenge, knowing that people fall prey to predictable errors and biases in judgment and decision-making, why think that the choice architect would be any less susceptible to these errors and biases when they are planning the choice environment? There are two tracks to take in response to this challenge. One is to argue that indeed, choice architects *are* less susceptible to errors and biases. The other is to argue that even with errors and biases, choice architecture designed to improve choices is better than the alternative, which is to allow choice architecture that makes people *worse off* go unchecked, or to allow people's choices to be influenced by random and unstructured choice environments.

Let us take the first track, that choice architects might be less susceptible to the errors and biases of ordinary decision-makers. Thaler and Sunstein nod towards this when they write, "…we stress that—as Adam Smith made clear—there are gains from specialization. It is

---

[14] The outcome of the veil of ignorance test might be different if iron consumption by hemachromatosis patients resulted in catastrophe (e.g., death) for that small number of people.

possible…to hire a competent expert to design a choice environment in which individuals have an easier time making good decisions."[15] In other words, those who extensively study judgment and decision-making (including the common pitfalls) will be less susceptible to these pitfalls themselves. Yet, this claim needs further support. The gains from specialization that Smith was referring to in his 1776 *An Inquiry into the Nature and Causes of the Wealth of Nations* involved specialization in the various parts of the process of making a pin. Specialization in how to make a part of a pin through study and repeated performance will no doubt make one good at making a pin. Study about decisional processes may or may not make one good at avoiding the pitfalls of them. Empirical evidence is needed. But, there are two helpful distinctions to be drawn in thinking about the danger of choice architects falling prey to biases and pitfalls when they structure choices. One is the distinction between choice architect as designer and choice architect as chooser, and the other is the distinction between choice architecture at a grand level and choice architecture at an individual day-to-day level. With respect to the first distinction, just because the choice architect is himself susceptible to making choices in his private life that are influenced by decisional heuristics and biases, this does not necessarily mean that when he helps other people make choices he is similarly susceptible. There has been a large amount of literature illustrating the differences between decision-making for oneself and decision-making for others.[16] When we make a decision for another person, or assist them with their decision, we are capable of a certain distance that we are not capable of with ourselves. Decision-making for or advising others often generates a critical attitude that is not always engendered when we decide for ourselves. For example, if I am choosing a car I am admittedly likely to fall prey to

---

[15] C. Sunstein and R. Thaler, Interview with Annika Mengisen, "From Push to Nudge: A Q&A with the Authors of the Latter" Freakonomics Blog-NYTimes.com 15 Sept. 2008.
[16] L. Kray and R. Gonzalez, "Differential Weighting in Choice Versus Advice: I'll Do This, You Do That," *Journal of Behavioral Decision Making*, 12 (1999), 207-217

what is called "the availability heuristic" and think that a Honda is the best car to buy since that is what I have seen all around my neighborhood (even though I have studied this heuristic and know we are biased towards what is most vivid and recent as opposed to statistics and evidence). If, however, I am helping my friend choose a car, I am going to warn him that he is probably unduly biased towards the Honda since that is what is "available" in his mind, and I will probably encourage him to look at the statistics (and I may even look them up for him). I think this difference in decision making for self and other is a common phenomenon. There is also an important difference between choice architects operating at a grand policy or design level and ones operating at an impromptu, individual, day-to day-level. The choice architects operating at the grand level are often engaged in months (or years) of planning and reflecting, and are likely in consultation with other people. There are many safeguards in place to ensure that the choice architect is not himself falling prey to cognitive and affective biases and heuristics.

The second track that one could take in response to concerns about the susceptibility of choice architects to errors and biases when planning choice environments is to argue that allowing them, in this biased state, to design choice environments that attempt to prompt people to make better choices is better than the alternative, which is to allow choice architecture that makes people *worse off* to go unchecked, or to allow people's choices to be influenced by random and unstructured choice environments that unintentionally direct them to bad choices. Much of our choice environment is structured by people who have no concern for our wellbeing, e.g., advertisers who subtly, creatively, and pervasively prompt us to buy their products, consume their goods, and adopt their way of life. In light of this, it would be preferable to have choice architects concerned with structuring choices reflectively and responsibly in a way that

makes people better off even if they are occasionally subject to biases and errors in judgment and decision-making.

D. Motivational Problems Such as Indifference, Conflicts of Interest, or Malevolence Affecting the Choice Architect's Judgment

One might grant that decisional biases and errors can be minimized in the choice architect, but point to the existence of motivational problems affecting the choice architect's ability to plan choice environments that make choosers better off. For example, choice architects could be indifferent, have conflicts of interest, or be malevolent. Salvat has criticized advocates of choice architecture for their failure to distinguish between the "functional" and the "personal" aspects of the choice architect. He refers to the example of the cafeteria planner choice architect who arranges the environment in which students make food choices, and questions her personal motivation to arrange the food in a way that makes students better off when all she needs to do to fulfill her function is to arrange the food in any which way. He writes, "She has fulfilled her professional responsibility. Ordering the dishes at random is not only possible but also worthwhile since it is the simplest and quickest…. Once she has fulfilled her function as a director of the cafeteria, she has no more reason a priori to be benevolent than to be malevolent towards her client."[17] The cafeteria planner is potentially indifferent about the well being of the students, or she has a conflict of interest in that it benefits her to minimize the time and effort that she puts into studying and constructing architecture that might make the students better off. Conflicts of interest in choice architects are going to be common. Even if chooser wellbeing is among their interests, they may also have self interests that conflict with chooser interests. For example, marketers and advertisers may or may not care about choosers' wellbeing, but they care

---

[17] Salvat, "Is Libertarian Paternalism an Oxymoron?" 11

significantly about their bottom line, and often that bottom line conflicts with the wellbeing of the choosers. Similarly, elected officials and policy makers may have conflicts of interest such as financial gains or the interests of particular private groups. Advocates of choice architecture recognize this. As Thaler and Sunstein concede, "…choice architects in all walks of life have incentives to nudge people in directions that benefit the architects (or their employers) rather than the users."[18] But, they rightly push detractors on what conclusions should be drawn from this observation. It is impossible to avoid any sort of structuring and influencing of people's choices even if unintentional.   The rules that are set, the way that choices are presented, the way that the environment is designed, etc. "…inevitably affect some choices and outcomes."[19] Instead, the focus should be on re-shifting incentives, monitoring, self-regulation, etc., to help offset any motivational issues in choice architects that may interfere with them designing environments that make other people better off.

One might grant that decisional biases and errors can be minimized in the choice architect, and that motivational problems can be controlled, but nonetheless have concerns about epistemological difficulties in discerning what really will make choosers better off. This is a legitimate concern, especially when it comes to certain types of choices. While it may not be difficult to gain knowledge of whether people would be better off if they avoided using high interest credit cards or over-consuming foods high in trans fats, it is difficult to know whether people would be better off if they took a year off between college and postgraduate study, had chocolate cake or strawberry shortcake, or had a lumpectomy or mastectomy in the case of breast cancer. In other words, some choices seem to be much more "preference sensitive" than others.

---

[18] Thaler and Sunstein, *Nudge*, p. 239
[19] *Ibid,* p. 11

A high degree of difference in individual preferences is especially common in the areas of religion, parenthood, and culture.[20]

While differences in preferences exist to be sure, there is much less variety than one might think, and much more data about what really does (and does not) make people better off. Repeated surveys have found that the following factors correlate with wellbeing (conceptualized as positive feelings and feelings of satisfaction over time): marriage, social relationships, employment, perceived health, religion, and the quality of government.[21] Moreover, we have learned that the health conditions of chronic pain, depression, and the onset of a fatal disease such as cancer, resist adaptation and take an enormous toll on wellbeing.[22] It is, however, essential that choice architects consult this data and not just rely on their intuitions about what would make people better off, because those intuitions might be wrong. For example, empirical research has shown that physicians are often highly inaccurate at predicting the goals and preferences of their patients.[23] This is in part due to the fact that physicians overweigh the value of health when they think about what will make people better off (as academics might overweigh education, or artists might overweigh creative expression, etc). Choice architects need to gain expertise in the choices that they are constructing and influencing; they need to know what the satisfaction and happiness levels are across the various possible choice outcomes.

There is one final point about epistemology to be brought up here, and that is the concern that choice architecture does not allow for gains in knowledge. Choosers typically learn by

---

[20] S. R. Ben-Porath, *Tough Choices: Structured Paternalism and the Landscape of Choice* (Princeton University Press, 2010), p. 10

[21] D. Bok, *The Politics of Happiness: What Government Can Learn From the New Research On Well-Being* (Princeton and Oxford: Princeton University Press, 2010), p. 17

[22] *Ibid,* p. 21

[23] J. Teno et al., "Preferences for Cardiopulmonary Resuscitation: Physician-Patient Agreement and Hospital Resource Use," *Journal of General Internal Medicine*, 10 (1995), 179-186

trying out different things and making mistakes, and since choice architecture nudges people directly to the best outcome, it does not seem to allow for this learning process. In response to this concern, I have three points. First, the process of coming to learn for oneself what makes one better off has its value, but that value can be outweighed by other things, such as the consequences of the mistakes made along the way and the benefits of achieving what makes one better off via a much shorter route. Second, it need not be the case that structuring choice environments *prevents* people from trying out different ones. Every now and then I might smoke a cigarette even though fortunately my usual choice environment leads me away from that (e.g., there are not cigarettes in the house, smoking is not allowed at work, I do not have any friends who smoke, etc). Third, it need not be the case that benevolently structured choice environments prod everyone to the *same* good. Return to the example of the choice architect in the cafeteria. While it is true that she is aiming to design a choice environment that prods the choosers towards healthy foods, there are a wide variety of healthy foods that the choice architect can prod towards—one day healthy Asian food, another day healthy Mexican food, etc. Choice architecture still allows for people to try different things and learn about their preferences.

I now want to turn to the second half of the paper, which addresses the other major claim of advocates of intentional choice architecture: that doing so is entirely compatible with individual liberty.

III. COMPATIBILITY WITH INDIVIDUAL LIBERTY

Those who advocate choice architecture claim that "The libertarian aspect of our strategies lies in the straightforward insistence that, in general, people should be free to do what they like—and to opt out of undesirable arrangements if they want to. To borrow a phrase from

the late Milton Freedman…people should be "free to choose"…. When we use the term libertarian…we simply mean liberty-preserving…. choices are not blocked, fenced off, or significantly burdened."[24] In order to be liberty preserving, choice environment designs, or interventions, must not block or significantly burden people's choices. This seems to be a necessary condition for liberty-preservation, but (1) more needs to be said about what it would *mean* to "significantly burden choices," and (2) other necessary conditions need to be added to this. Let us first discuss the other necessary conditions that need to be added.

A. The Problem of Manipulation and the Transparency Requirement

The requirement that a choice environment intervention not block or significantly burden choices is not by itself sufficient for it to be liberty preserving. Let me be more precise. The requirement that the intervention not block or burden choice is not sufficient for it to be *autonomy* preserving, and autonomy is something that those concerned with freedom of choice and action are typically concerned with. Gerald Dworkin has distinguished between liberty and autonomy, arguing that liberty is "…the ability of a person to do what he wishes and to have significant options that are not closed or made less eligible by the actions of other agents or the workings of social institutions," whereas autonomy is "…understood as a power of self-determination."[25] Dworkin gives Locke's example of a man who is put into a prison cell and told that it is locked when it really is not as an example of a person with liberty but not autonomy. The prisoner lacks autonomy because "His view of the alternatives open to him has been manipulated by the guards in such a fashion that he will choose not to leave."[26] The prisoner's autonomy is impaired "…if we turn to a notion of autonomy that involves people being able to

---

[24] Thaler and Sunstein, *Nudge*, p. 5
[25] Dworkin, *The Theory and Practice of Autonomy,* (Cambridge University Press, 1988), p. 105
[26] *Ibid.*

shape their own lives in important ways, not being subject to the influence and power of others….”[27]

Those who care about freedom of choice and action are concerned not just with liberty narrowly conceived but also with autonomy. As Sigal Ben-Porath has argued, in order for a choice to be properly available and successfully achieved, three components are necessary: (1) autonomy (the ability to discern and consider options, and the capability to act according to one's preferences), (2) freedom (political structure that allows for the realization of rights), and (3) social opportunity (the availability of multiple relevant options for the individual to choose from).[28] The crucial question then is which types of choice architecture interfere with a person's self-determination or autonomy. Dworkin distinguishes between acting for causes (e.g., someone runs up to a person and yells “JUMP!” in their ear) and acting for reasons (e.g., someone challenges the person to jump), and he further distinguishes between creating reasons for action versus calling attention to reasons that already exist, between processes that bypass any appeal to reason versus processes that work via reasons, and cases where reasons appealed to are made explicit versus used covertly. Instead of putting forth a sharp division regarding which of these cases interfere with autonomy and which do not, Dworkin suggests “In order to understand why acting for certain reasons is interference with freedom and others not, one cannot just look at the actual situation in which the agent acts.” One has to ask “Does he mind acting for these particular reasons or not?”[29] This hierarchical, reflective, historical view of

---

[27] *Ibid,* p. 164
[28] Ben-Porath, *Tough Choices*, p. 10
[29] Dworkin, *The Theory and Practice of Autonomy*, p. 155

autonomy is one that has been adopted, in various forms, by Frankfurt, Christman, and Bratman, among others.[30]

The crucial point is that even if choice architecture does not block or significantly burden choices, it might still interfere with a person's ability to discern and consider options and act according to her own preferences, i.e., it might interfere with her autonomy. Those who are concerned with freedom of choice and action are also concerned about autonomy, and thus it is not sufficient that choice architects not block or burden choices. Choice architects must also ensure that agents retain the ability to discern and consider options and the capability to act according to their own preferences.

Let us look at an example. Jane is an unhealthy eater. Her roommate is concerned about the amount of junk food that she eats and decides to create a choice environment that nudges her to make healthier eating choices. Jane's roommate does this by replacing the normal mirrors in their apartment with distorting mirrors that make Jane look overweight.[31] Jane's roommate does not engage in rational argument with her, nor does she threaten her. Instead, she manipulates her, i.e., she bypasses her reasoning by exploiting non-rational elements of psychological processing and/or by influencing choices in a way that is not obvious.[32] The roommate's choice

---

[30] J. S. Taylor, "Introduction" in J. S. Taylor (ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, (Cambridge University Press, 2005), pp. 1-29
[31] Thaler and Sunstein, *Nudge*, p. 244
[32] A. Mele, *Agents: From Self-Control to Autonomy*, (New York, NY: Oxford University Press, 2001), p. 145. There are various ways in which manipulation can occur. Examples can include imparting false information, leaving out information, getting people to look at red herrings to provide irrelevant inputs or to crowd out relevant inputs, making insinuations, raising suspicions, using vivid imagery or loaded language, using leading questions or certain tones of voice, browbeating, and taking advantage of emotional needs and fears. For more on manipulation see R. Noggle, "Manipulative Actions: A Conceptual and Moral Analysis," *American Philosophical Quarterly*, 33 (1996), 43-55; M. Baron, "Manipulativeness," *Proceedings and Addresses of the American Philosophical Association*, 77 (2003), 37-54; P. Greenspan, "The Problem With Manipulation," *American Philosophical Quarterly*, 40 (2003) 155-164; and E. Cave, "What's Wrong With Motive Manipulation," *Ethical Theory and Moral Practice*, 10 (2007) , 141.

environment intervention does not block or significantly burden Jane's choice to eat a lot of junk food, but does it (1) affect her ability to discern her options, consider them, and act in accordance with her own preferences, or (2) cause her to act for reasons that she would mind acting for?   If so, then the intervention poses a threat to her freedom of choice and action. I suspect that in this case Jane would object to these reasons for acting if it were disclosed to her that her reasons for eating healthy emerged from her roommate's actions. As Dworkin notes of the classic gunman example where a gunman threatens to shoot a victim if he does not hand over his money, "People resent acting merely in order to retain a status quo against the interference of another agent (threats)."[33] Jane likely resents acting for her roommate's deceptive and manipulative actions. I do not mean to suggest that all cases of manipulation, or of choice architecture, necessarily impair autonomy (or that the ones that do, do so to the same extent), but they certainly raise red flags that are worth investigation.[34]

One way to combat choice architecture hindering the exercise of autonomy is to make it transparent to those who encounter it. In fact, Thaler and Sunstein do toy with the idea of transparency. In order to express the transparency requirement they draw on Rawls's "publicity principle," which would ban interventions that choice architects would not be able or willing to defend publicly.[35] The problem with this requirement is that it is only hypothetical. Certainly, considering whether one would hypothetically be able or willing to defend a choice environment

---

[33] Dworkin, *The Theory and Practice of Autonomy*, p. 155

[34] Hausman and Welch do think that all cases of "nudging" that involve "taking advantage of people's psychological foibles" impair autonomy. See D. Hausman and B. Welch, "Debate: To Nudge or Not to Nudge," *The Journal of Political Philosophy*, 18 (2010), 128, 133. Although, from this they extrapolate that the use of decisional psychology techniques to shape people's decisions should be avoided and that rational persuasion (the use of data and valid arguments) is the only defensible alternative. Yet, they repeatedly acknowledge the difficulty and potential impossibility of the use of rational persuasion (data and valid arguments) alone, acknowledging that what may be doing the persuasive work is the accompanying soft smile, aura of authority, order in which the data is presented, etc. Thus, I find their position untenable since it merely turns a blind eye to the facts and promotes denial instead of responsible reflection on and management of the factors that shape people's choices.

[35] Thaler and Sunstein, Nudge, p. 244

design offers some normative guidance for the choice architect, but it is not sufficient to preserve freedom of choice and action. In order to preserve freedom of choice and action, the transparency would need to be actual not merely hypothetical. How this transparency requirement would be fulfilled would vary from case to case. In some cases, the choice architecture and its intention will be readily transparent and nothing would need to be disclosed, in other cases a verbal disclosure would be best, and in some cases a posted written disclosure would be best.

But of course, disclosure and transparency is not enough to preserve liberty since manipulation (which at least poses a threat to a person's ability to govern themselves) occurs not just when a person is influenced without knowing it, but when they are influenced via exploitation of the non-rational elements of their psychological make-up. Return to the example of Jane's roommate who places unflattering mirrors around the apartment. Even if the roommate transparently *tells* Jane what she is trying to do, by using the mirrors in her choice environment she is exploiting her desire to feel attractive, her guilt, etc. Part of what is threatening to Jane's liberty here, as is true with many cases of manipulation via exploitation of emotion, is that Jane's roommate is making it very hard for Jane "go her own way" and eat the junk food.

B. The Problem of Blocking or Burdening Choices

The notion that we must avoid making it hard for a person to discern her options and act according to her preferences is essential for the preservation of liberty. In order to fully preserve liberty choice architecture must not block off or significantly burden other choices. Of course, developing criteria for when choice architecture is "easily" overcome or avoided is tricky, but I will attempt to make a few guiding remarks here. In an attempt to deal with this challenge,

Thaler and Sunstein write, "We do not have a clear definition of "easily avoided,"…. It would be arbitrary and a bit ridiculous to offer an inflexible rule to specify when costs are high enough to disqualify a policy as libertarian, but the precise question of degree is not really important. Let us simply say that we want those costs to be small."[36]

It is true that it would be arbitrary to offer such an inflexible rule, simply because liberty, freedom, and autonomy are concepts best conceptualized as degree concepts that lie on a spectrum. There are, however, a couple of points to make. First, the "costs" of avoiding or overcoming an instance of influence via choice architecture can take many different forms. For example, a person may have to give up time, take on an annoyance, walk a little further, pay a little more money or forfeit some money or material goods, execute willpower, feel bad about himself (self-scorn), undergo social scorn, or have to think a little harder. The second point is that the cost of willpower is particularly important since it threatens a person's ability to govern themselves the way that they want to govern themselves; it threatens their autonomy and as such their freedom of choice and action.

Consider an example of a woman, Susan, who wants her colleague David to become romantically involved with her, but David does not want to become involved with her, so Susan exploits his weakness for the scent of lavender and designs a choice environment that involves a lot of lavender. Susan is not blocking off any options for David, and it is not as if David has to pay a substantial amount of money or forfeit a lot of his time in order to overcome Susan's architecture, but it does require a lot of willpower. The fact that extensive willpower is necessary makes it difficult for him to govern himself the way that he wants to. David's

---

[36] *Ibid*, p. 251. Interestingly, while here the adjective that they use for the costs is 'small,' at other times they use 'close to zero."

autonomy is threatened despite the fact that Susan's attempts are transparent, and it is threatened because the architecture requires that David muster a lot of willpower making it difficult for him to go his own way.

To sum up then, in order for an instance of choice architecture to be liberty-preserving in a meaningful way, choice architects must ensure that (1) other choices are not blocked off, (2) avoiding or overcoming the influence attempt is easy and can be done with low material, physical, temporal, mental, and willpower costs, and (3) influence attempts are transparent and/or that the person influenced would not mind acting for reasons of the choice architecture. It truly must be easy for a person to go her own way, to govern herself.[37]

C. What Would a True Libertarian Do?

Some might object that a "true libertarian" would not engage in choice architecture at all. Instead, they would favor ways to help people make better decisions that do not involve architecting their decisions for them. One way that is often mentioned is to use "mandated choice," whereby instead of influencing people towards a particular choice, we force them to make their own choice. The second way is to "debias" people instead of "rebiasing" them. In other words, one might recognize that human decision making is full of biases and errors that result in less than optimal and sometimes harmful decisions, but hold that the response should be to strip people of these harmful biases instead of rebiasing them towards better decisions. I will take each point in turn. Regarding the option of mandated choice, I have two responses. One is that often people do not want to choose. Choosing requires a lot of time, energy, and

---

[37] An interesting side issue about liberty that I will raise, though not address in this paper, is the issue about the freedom of institutions/companies to choose *not* to engage in choice architecture to promote people's well-being and the overarching regulatory issues that might be raised by institutions/companies that choose not to engage in choice architecture for the good of their consumers or clients (e.g., the supermarket manager who refuses to put the fruits and vegetables near the front of the store and the unhealthy goods in the back).

psychological burden—not to mention that the person must enhance their expertise in the area that they are making a choice about. For example, policy makers could force me to make a choice about which retirement plan I join, but I would be none too happy about the time and energy that I would have to spend investigating all of the options. As Thaler and Sunstein note, "Given that people would often choose not to choose, it is hard to see why freedom lovers should compel choice even through people (freely and voluntarily) resist it."[38] And of course, those who do want to investigate the options and make the choice themselves can certainly do so. The second point that I want to make regarding mandated choice is that it does not avoid influencing people's choices. Imagine that someone is describing retirement plans to me and instead of setting a default such that I am enrolled in the plan that suits me best; they force me to make a choice. The person still has to describe the options to me; and my choice will be unavoidably influenced by the order and manner in which they present the options. So for the libertarian, mandated choice is not a realistic or better alternative to choice architecture.

What about the approach of identifying the biases that are resulting in bad decisions and trying to remove them? Why not encourage people to use their reflective systems more in decision making instead of setting up things so that their continued use of their automatic systems will result in better outcomes? Why not just inform people and educate them? Mitchell is an advocate of this "debiasing approach" and cites several studies showing that asking experimental subjects to consider alternate or opposing options or arguments ameliorated the effects of several biases.[39] I have two responses to this approach. One is that even if you work to rid someone of one of their biases and do so successfully, other biases are still present and new ones have likely been introduced in the process. It is simply impossible to completely debias

---

[38] Thaler and Sunstein, *Nudge*, p. 243
[39] Mitchell, "Libertarian Paternalism Is an Oxymoron" 14

someone and make them into a rational, robotic, calculating decision-maker. It is impossible to avoid influencing people's choices. It is impossible to be neutral "no matter how scrupulously designers try to achieve that goal."[40] This is the key lesson from behavioral psychology, behavioral economics, and decision science. Second, even if one could debias people via the presentation of [neutral] information, education, and engagement of the reflective system, this approach is simply not practical and as such is not desirable. Much of decision-making simply does not have the opportunity for this sort of extensive intervention. We make dozens of decisions every day: how fast to drive, whether to stop at that yellow light, whether to buckle our seat belts, whether to work or check facebook, whether to eat pizza or salad, whether to take the stairs or the elevator, whether to work out or turn on the TV, whether to go out or eat in, what to buy at the grocery store, and whether to leave the water running while we brush our teeth. If we want to improve people's decisions in these areas then we must interact with their automatic systems. There is not time to ask them to consider all of the options and the arguments on the other side. Someone might urge that for the more important, less mundane, less day-to-day decisions, the "debias/engage the reflective system" rather than the "rebias/architect the automatic system" approach be implemented. This is a fair point, but we must remind ourselves that the effort to debias will never be fully successful and moreover that it may not be as effective as fashioning the automatic system to encourage people to make choices that make them better off. To see this latter point, imagine a person deciding whether to undergo a surgery. They are very much influenced by what is automatically available in their mind, which is the image of their grandmother who underwent a similar surgery and died. A choice architect could encourage the person to really consider the statistics, which point very favorably towards undergoing surgery, and to engage in constructing arguments on the opposite side; but this is not

---

[40] Thaler and Sunstein, *Nudge*, p. 243

going to be as moving as placing a vivid example in their mind that will be automatically available of someone who died without the surgery. In other words, rebiasing through crafting the automatic system will often be the most effective means towards our end of helping people make better choices.

Before concluding there is one final issue that I would like to take up and that is the claim that choice architecture and its effects cannot be avoided, that influencing choices is impossible to avoid. One might grant this point but maintain that there is important commission/omission distinction that a "true libertarian" would uphold, and that is that there is a difference between being influenced by a random environmental "other" and another person intentionally designing the influential environment. The former is a case of influence by omission (we do not do anything except for allow the environment, broadly construed, to influence people) and the latter is a case of influence by commission (we actively do something to the environment so that it influences people in a particular way). For those concerned about freedom, it might seem like there is something particularly objectionable about designed environmental influence *by another person*. Dworkin raises this point, although does not flesh it out, when he notes the difference between a natural world/culture and a planned one (such as the world in Walden Two).[41] Frankfurt, on the other hand, views the difference to be irrelevant. He writes, "We are inevitably fashioned and sustained, after all, by circumstances over which we have no control…It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberatively manipulative design of other agents."[42] The crucial question is *whether* I am in control and whether I have liberty. If I do not,

---

[41] Dworkin, *The Theory and Practice of Autonomy*, p. 152
[42] H. Frankfurt, "Reply to John Martin Fischer," in S. Buss and L. Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, (MIT Press, 2002), 27-31

whether because of a gun to my head or a mental illness or an instance of choice architecture then I do not. The question of whether an agent is in control and at liberty is best answered by asking questions about ease and cost of going one's own way, transparency, and whether one minds acting for the reasons that she did.

I will end the paper with a quote from Thaler and Sunstein, along with a caveat and a point of emphasis. Thaler and Sunstein write, "Choice architecture and its effects cannot be avoided, so the…answer is an obvious one, call it the golden rule…offer nudges that are most likely to help and least likely to inflict harm."[43] As caveat to that I would add "And least likely to disrespect autonomy." The point of emphasis is that judgments about what is "most likely to help" or "make people better off" should be based on data about satisfaction and happiness levels across various outcomes and not simply on intuition. Choice architecture can indeed be a mechanism to improve people's choices while preserving their liberty, although as this paper shows, there are many stipulations, nuances, and clarifications involved in establishing the soundness of that claim.

---

[43] Thaler and Sunstein, *Nudge*, p. 72