



Freedom: The Holberg Lecture, 2018

Citation

Cass R. Sunstein, Freedom: The Holberg Lecture, 2018 (June 2, 2018).

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:37292812>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

Freedom: The Holberg Lecture, 2018

Cass R. Sunstein*

Abstract

*If people have freedom of choice, do their lives go better? Under what conditions? By what criteria? Consider three distinct problems. (1) In countless situations, human beings face a serious problem of “navigability”; they do not know how to get to their preferred destination, whether the issue involves health, education, employment, or well-being in general. This problem is especially challenging for people who live under conditions of severe deprivation, but it can be significant for all of us. (2) Many of us face problems of self-control, and our decisions today endanger our own future. What we want, right now, hurts us, next year. (3) In some cases, we would actually be happy or well-off with two or more different outcomes, whether the issue involves our jobs, our diets, our city, or even our friends and partners, and the real question, on which good answers are increasingly available, is what most promotes our welfare. The evaluative problem, in such cases, is especially challenging if a decision would alter people’s identity, values, or character. Private and public institutions -- including small companies, large companies, governments -- can help people to have better lives, given (1), (2), and (3). This Essay, the text of the Holberg Lecture 2018, is the basis for a different, thicker, and more elaborate treatment in a book, *On Freedom* (forthcoming, Princeton University Press, 2019).*

It is true, but far too little, to say that I am humbled to have been awarded the Holberg Prize for 2018. The Prize covers a dauntingly large territory: “the arts and humanities, social

* Robert Walmsley University Professor, Harvard University. This lecture is the final, unaltered version of the Holberg Lecture 2018, delivered on June 5, 2018 at the University of Bergen. I am grateful to my hosts at the Holberg Committee and the University of Bergen for their extraordinary generosity and kindness on that occasion. Special thanks to Ellen Mortenson, Ole Andreas Sandmo, and Solveigh Stornes. The text is the basis for a lengthier treatment in a book, *On Freedom*, forthcoming from Princeton University Press in 2019.

I am grateful to Jacob Goldin, Stephen Greenblatt, L.A. Paul, and Lucia Reisch for valuable comments on earlier versions. I am also grateful to Madeleine Joseph for excellent comments and research assistance. Some of these ideas were presented to the Department of Philosophy at the University of North Carolina, Chapel Hill, where I received exceptionally valuable help. Special thanks to Richard Thaler, my collaborator on all things related to nudges and nudging. This lecture draws, in some places, on Cass R. Sunstein, “*Better Off, as Judged by Themselves*”: *A Comment on Evaluating Nudges*, 64 INT’L REV. ECON. 1 (2018).

science, law or theology.” My principal focus is law, but I studied literature in college, and it made a large mark on me. In the last decades, social sciences, including psychology and behavioral sciences, have been a major interest of mine. I have been lucky enough to have learned from all of the fields covered by the Holberg Prize.

My topic here has preoccupied artists and poets and novelists, psychologists and anthropologists and philosophers, lawyers and theologians: the relationship between freedom of choice and human welfare.

I have two epigraphs to offer on that topic. The first should be familiar:

So when the woman saw that the tree was good for food, and that it was a delight to the eyes, and that the tree was to be desired to make one wise, she took of its fruit and ate; and she also gave some to her husband, who was with her, and he ate.

Then the eyes of both were opened, and they knew that they were naked; and they sewed fig leaves together and made loincloths for themselves.¹

The second epigraph is from my favorite literary text on the topic of freedom, A.S. Byatt’s *Possession*, after a fateful choice (and yes, it involved a romance):

In the morning, the whole world had a strange new smell. It was the smell of the aftermath, a green smell, a smell of shredded leaves and oozing resin, of crushed wood and splashed sap, a tart smell, which bore some relation to the smell of bitten apples. It was the smell of death and destruction, and it smelled fresh and lively and hopeful.²

Byatt is speaking of a free choice, and a kind of fall, and her tale overlaps with that of Genesis. Byatt’s account is more upbeat. Let us be clear: Every human being is blessed to experience that smell.

In its broadest form, my questions are whether and when people’s free choices will make their lives go better. The liberal political tradition often offers a simple answer to that question: *Yes, and usually.* Artists, novelists, psychologists, and theologians tend to disagree, and they are right to insist that those answers are far too simple.

In brief: People might not know how to get where they want to go. They can be tempted. Background conditions greatly matter. Sometimes people lack self-control. Sometimes people’s choices are not, in the deepest sense, their own; they are deprived, deceived, or manipulated. Sometimes people lack crucial information. Sometimes their preferences are a product of injustice or deprivation. Sometimes they simply blunder. As a result, their lives go much worse.

¹ *Genesis* 3:6. A magnificent treatment is STEPHEN GREENBLATT, *THE RISE AND FALL OF ADAM AND EVE* (2017).

² A.S. BYATT, *POSSESSION* 551 (1990).

To make progress, I will focus largely on freedom-preserving interventions, sometimes called “nudges,” which steer people in certain directions, but which also allow people to go in the direction and toward the destination that they prefer. But I shall have something to say about coercion as well. One of my central goals is to put a bright spotlight on the problem of *navigability*, and to suggest that we have paid far too little attention to it. All over the world, people suffer because life is not easily navigable. Those who are facing hardship — poverty, mental illness, chronic pain — are often unable to solve the problem of navigability. They deserve some help.

I will be arguing that in the coming years, those who are interested in freedom, welfare, and uses of behavioral science should devote far more attention to navigability, writ very large. More ambitiously, I will be arguing for the importance of exploring the importance and limits of freedom by investigating the ingredients of human welfare, in a way that is heavily empirical — and of using what we learn for purposes of policy.

Orientation

For orientation: In daily life, a GPS device is an example of a nudge. It respects freedom of choice; you can ignore its advice if you like. But it helps you to get where you want to go. Other nudges include “apps” that tell people how many calories they ate during the previous day; text messages, informing customers that a bill is due or that a doctor’s appointment is scheduled for the next day; automatic enrollment in pension plans; the default settings on computers and cell phones; systems for automatic payment of credit card bills and mortgages.

In government, nudges include graphic warnings for cigarettes; mandatory labels for energy efficiency or fuel economy; “nutrition facts” panels on food; the “Food Plate,” which provides a simple guide for healthy eating; default rules for public assistance programs; a website like data.gov or data.gov.uk, which makes a large number of data sets available to the public.

Many policies, by contrast, take the form of *mandates* and *bans*. For example, the criminal law forbids theft and assault. It does not allow people to go their own way. Other policies take the form of *economic incentives* (including disincentives), such as subsidies for renewable fuels, fees for engaging in certain activities, or taxes on gasoline and tobacco products. These policies play important roles, but they are not nudges. Mandates and bans eliminate freedom of choice. Incentives do not, but they impose a kind of skew. If nudges do that as well, it is not because they impose material burdens or grant material benefits.

A primary goal of behaviorally informed policy has been to insist that many interventions, both private and public, decisively affect choices and outcomes, even if they do not alter incentives. Because of the appeal of choice-preserving approaches, many nations have shown keen interest in nudges and nudging, and have established Nudge Units, or Behavioral Insights Teams, to increase health and safety or otherwise to promote important goals. By any measure, the consequences of some nudges are not properly described as modest — in the United States, the United Kingdom, the Netherlands, Ireland, Denmark, Singapore, and numerous other nations. As a result of automatic enrollment in free school meals programs, more than 11 million

poor American children are now receiving free breakfast and lunch during the school year. Credit card legislation, enacted in 2010, is saving consumers more than \$10 billion annually; significant portions of those savings come from nudges and nudge-like interventions.³ With respect to savings, automatic enrollment in pension programs has produced massive increases in participation rates.⁴

New nudges, now in early stages or under discussion, could also have a major impact. If the goal is to reduce greenhouse gas emissions, automatic enrollment in green energy can have large effects. The Earned Income Tax Credit is probably the most effective anti-poverty program in the world, but many eligible people do not take advantage of it. Automatic enrollment would have large consequences for the lives of millions of people. In many nations, automatic voter registration could turn millions of people into voters.

With respect to the world's most serious problems, the use of nudges remains in its preliminary stages. We will see far more in the future, and the impact will not be small.

Inevitability

Here is 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?'"⁵

This is a tale about choice architecture — the background against which choices are made. Choice architecture is inevitable, whether or not we see it, and it nudges us. It is the equivalent of water. Weather is itself a form of choice architecture, because it influences what people decide; on snowy days, people are especially likely to purchase cars with four-wheel drive, which they will return to the market unusually quickly.⁶ Human beings cannot live without some kind of weather. Nature nudges. The law of contract, tort, and property is a regulatory system, and it will nudge, even if it allows people to have a great deal of flexibility.

In this light, choice architecture, and some kinds of nudging, are inevitable. Human beings cannot wish them away. Any store, real or online, must have a design; some products are seen first, and others are not. Any menu places options at various locations. Television stations are assigned different numbers, and strikingly, the number matters, even when the costs of

³ Sumit Agarwal et al., *Regulating Consumer Financial Products: Evidence from Credit Cards*, 130 Q.J. ECON. 111 (2014).

⁴ Raj Chetty et al., *Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark*, 129 Q.J. ECON. 1141 (2014); Richard H. Thaler, *Much Ado About Nudging*, BEHAV. PUB. POL. BLOG (June 2, 2017), <https://bppblog.com/2017/06/02/much-ado-about-nudging/>.

⁵ David Foster Wallace, *In His Own Words*, ECONOMIST: 1843 MAG., <https://www.1843magazine.com/story/david-foster-wallace-in-his-own-words>.

⁶ Meghan R. Busse et al., *Projection Bias in the Car and Housing Markets* (Nat'l Bureau of Econ. Research, Working Paper No. 18212, 2014), <http://www.nber.org/papers/w18212>.

switching are vanishingly low; people tend to choose stations with lower numbers.⁷ Any website has a design, which will affect what and whether people will choose.⁸

It would be possible, of course, to define choice architecture in a narrower way, and to limit it to intentional designs by human beings. There is nothing intrinsic to human language that rules out that definition, and for most purposes, I will be emphasizing intentional design. But if the goal is to see how and when people are influenced, the broader definition is preferable. It shows that our choices are often an artifact of an architecture for which no human being may have responsibility — a sunny day, an unexpected chill, a gust of wind, a steep hill, a full (and romantic) moon.

“As Judged by Themselves”

Some nudges are designed to reduce externalities; consider fuel economy labels that draw attention to environmental consequences, or default rules that automatically enroll people in green energy. But many nudges are designed *to increase the likelihood that people’s choices will improve their own welfare*. Richard Thaler and I argue that the central goal of nudging is to “make choosers better off, *as judged by themselves*.”⁹ I will be focusing much of my discussion here on that idea, and attempting to explain its complexities. As we shall see, navigability is an important part of the picture — but however important, it is only one part.

Social planners — or in our terminology, choice architects — might well have their own ideas about what would make choosers better off, but in general, the lodestar is people’s own judgments. To be a bit more specific: The lodestar is welfare, and people’s own judgments are a reasonable (if imperfect) way to test the question whether nudges are increasing their welfare.

It should be clear that this claim builds on the view, associated with John Stuart Mill, that individuals are in a unique position to know what will improve their welfare, and that outsiders will often blunder. Mill insists that the individual “is the person most interested in his own well-being,” and the “ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by any one else.”¹⁰ When society seeks to overrule the individual’s judgment, it does so on the basis of “general presumptions,” and these “may be altogether wrong, and even if right, are as likely as not to be misapplied to individual cases.” If the goal is to ensure that people’s lives go well, Mill concludes that the best solution is for public officials to allow people to find their own path. Consider in the same vein F.A. Hayek’s remarkable suggestion

⁷ See Gregory Martin and Ali Yurukoglu, *Bias in Cable News: Real Effects and Polarization* (Nat’l Bureau of Econ. Research, Working Paper No. 20798, 2014), <http://www.nber.org/papers/w20798>.

⁸ See STEVE KRUG, *DON’T MAKE ME THINK REVISITED: A COMMON SENSE APPROACH TO WEB AND MOBILE USABILITY* (2014).

⁹ RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* 5 (2009) (*italics in original*).

¹⁰ JOHN STUART MILL, *ON LIBERTY* 8 (Kathy Casey ed., 2002) (1859).

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*.”¹¹

Three Objections

It should be clear that even if this suggestion is correct, the “as judged by themselves” criterion runs into three immediate objections.

First: Suppose that people are manipulated into having some belief or engaging in some action — for example, buying a product or supporting a candidate. To evaluate this concern, we would need to define manipulation and specify what is wrong with it.¹² Let us stipulate, very briefly, that manipulation consists of a failure sufficiently to engage people’s capacities for reflection and deliberation; it is a way of tricking people. One problem with manipulation, thus defined, is that it does not treat people with respect. It is disrespectful of human agency. And because manipulators do not allow people to make choices about what will promote their own interests, and reflect their own values, they threaten to reduce welfare as well.

When people are manipulated, they are deprived of the (full) ability to make choices on their own, simply because they are not given a fair or adequate chance to weigh all variables. Victims of manipulations are less free. At the same time, manipulators often lack relevant knowledge — about the chooser’s situation, tastes, and values. Lacking that knowledge, they nonetheless subvert the process by which choosers make their own decisions about what is best for them. Things are even worse if the manipulator is focused on his own interests rather than on those of choosers. It is in this sense that a self-interested manipulator can be said to be stealing from people — both limiting their agency and moving their resources in the preferred direction. Manipulators are thieves. If manipulation has occurred and if it is wrong, we would have an objection, even if people are satisfied *ex post*.

Second: Suppose that people are nudged to accept a view or to engage in some act that runs into an independent moral objection. For example, people might be nudged to be racist or sexist. If the nudge works, they might be satisfied with their racism or sexism. The “as judged by themselves” criterion would be met. Even if so, the independent objection holds.¹³ To that extent, the criterion is not sufficient.

Third: Consider the chilling last lines of George Orwell’s *1984*: “He had won the victory over himself. He loved Big Brother.”

¹¹ FRIEDRICH HAYEK, *The Market and Other Orders*, in *THE COLLECTED WORKS OF F. A. HAYEK* 384 (Bruce Caldwell ed., 2013).

¹² For a detailed discussion, see CASS R. SUNSTEIN, *THE ETHICS OF INFLUENCE* (2016). For examples and discussion, see GEORGE AKERLOF AND ROBERT SHILLER, *PHISHING FOR PHOOLS* (2015).

¹³ We might also wonder about cases of diminished capacities. Suppose that people’s mental or physical capacities are reduced, but they believe themselves to be better off for one or another reason. Their own judgment might not be authoritative. I do not discuss this problem in detail, because it is hard to identify nudges that diminish people’s capacities.

These lines signal the final defeat of Orwell's hero, Winston Smith, who has been nudged (among other things) by Orwell's villain, O'Brien, not only to *do* as the Party wishes but to *think* as the Party wishes – and to embrace his new thinking (almost erotically; actually, we can delete the “almost”). It is true that Smith has not merely been nudged; he has been manipulated and terrified (and effectively coerced). But even if he had been nudged, an objection would remain. Smith ends up in a state not merely of unfreedom but of slavery, even if or perhaps because he is content with that state. So much, you might say, for the “as judged by themselves” criterion.

Aldous Huxley's *Brave New World* is a different kind of dystopian novel, but Huxley makes the same point: “A really efficient totalitarian state would be one in which the all-powerful executive of political bosses and their army of managers control a population of slaves who do not have to be coerced, because they love their servitude.”¹⁴ Hence the plea of Huxley's hero, the Savage: “But I don't want comfort. I want God, I want poetry, I want real danger, I want freedom, I want goodness. I want sin.”¹⁵ Or consider this passage¹⁶:

“All right then,” said the Savage defiantly, I'm claiming the right to be unhappy.”

“Not to mention the right to grow old and ugly and impotent; the right to have syphilis and cancer; the right to have too little to eat, the right to be lousy; the right to live in constant apprehension of what may happen tomorrow; the right to catch typhoid; the right to be tortured by unspeakable pains of every kind.”

There was a long silence.

“I claim them all,” said the Savage at last.”

What must be said here is that even if people believe that they are better off, and even if the “as judged by themselves” criterion is met, they might be in a state of unfreedom, because they might not be experiencing a sufficiently human life, in which they are able to enjoy agency and autonomy. That is a fundamental point. But it should not be taken as an objection to nearly all real-world nudging, at least in democratic societies, where the goal is often to increase people's capacity for agency and autonomy (consider disclosure of information), and when efforts to make life easier hardly create servitude (consider automatic enrollment in pension plans).

Ambiguities

In this light, we might conclude that while it is (almost always) necessary to meet the “as judged by themselves” criterion, meeting that criterion is not (always) sufficient. The conclusion is generally right, though it needs to be specified. And beyond these qualifications, it is certainly reasonable to wonder about potential ambiguities in that criterion. Most important:

¹⁴ ALDOUS HUXLEY, *BRAVE NEW WORLD* xii (1931).

¹⁵ *Id.* at 163.

¹⁶ *Id.*

- Should the focus be on choosers' judgments before the nudge, or instead after? What if the nudge alters people's preferences, so that they like the outcome produced by the nudge, when they would not have sought that outcome in advance? What if preferences are constructed by the relevant choice architecture?
- What if people's ex ante judgments are wrong, in the sense that a nudge would improve their welfare, even though they do not think that it will?
- Do we want to ask about choosers' actual, potentially uninformed or behaviorally biased judgments, or are we entitled to ask what choosers would think if they had all relevant information and were unaffected by relevant biases?

To make progress on such questions, three categories of cases should be distinguished:

(1) those in which choosers have clear antecedent preferences, and nudges help them to satisfy those preferences;

(2) those in which choosers face a self-control problem, and nudges help them to overcome that problem;

(3) those in which choosers would be content with the outcomes produced by two or more nudges, or in which ex post preferences are endogenous to or constructed by nudges, so that the "as judged by themselves" criterion leaves choice architects with several options, without specifying which one to choose.

Cases that fall in category (1) plainly satisfy the "as judged by themselves" criterion — and there are many such cases. From the standpoint of the "as judged by themselves" criterion, cases that fall in category (2) — reminiscent of *Genesis and Possession* — are also unobjectionable, indeed they can be seen as a subset of category (1); and they are plentiful. But as the passage from Byatt suggests, they have their own complexities. Cases that fall in category (3) create special challenges, which may lead us to make direct inquiries into people's welfare or to explore what informed, active choosers typically select.

Navigability, Writ Large

I have already signaled one of the central points that I mean to press: As the GPS example suggests, many nudges and many forms of choice architecture 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 helpful choice architecture is desirable as a way of promoting simple navigation.

Making the world more navigable might not seem like the most ambitious imaginable idea, but it deserves far more attention. Many of the problems faced by human beings stem from

insufficient navigability. I have noted that for poor people, that is indeed a pervasive problem. Consider these words from Esther Duflo¹⁷:

We tend to be patronizing about the poor in a very specific sense, which is that we tend to think, ‘Why don’t they take more responsibility for their lives?’ And what we are forgetting is that the richer you are the less responsibility you need to take for your own life because everything is taken care for you. And the poorer you are the more you have to be responsible for everything about your life. . . . [S]top berating people for not being responsible and start to think of ways instead of providing the poor with the luxury that we all have, which is that a lot of decisions are taken for us. If we do nothing, we are on the right track. For most of the poor, if they do nothing, they are on the wrong track.

In my terms here, the problem is that they have to *find* the right track — to identify the right doctor, to find the right job, to get help in taking care of children. With respect to freedom, that is a serious problem. All over the world, efforts to increase navigability can make all the difference. Good cities are easily navigated; so are good airports; so are good hotels; so are good websites. We might think of efforts to increase navigability as a form of *means paternalism*. A GPS device respects people’s ends; it does not quarrel with their judgment about their preferred destination. But it helps them how to get where they want to go. Many interventions, having nothing to with literal navigation, can be understood in similar terms, including default rules, information, reminders, and warnings.

The stakes can be high. In recent years, a great deal of attention has been focused on the idea of happiness, which we may take, for present purposes, as a surrogate for welfare. In many nations, unhappiness is a product of mental illness. In many others, it is a product of unemployment. Increases in navigability can ensure that those who suffer from mental illness to get some help. Increases in navigability can help people to find employment. With such approaches, freedom is hardly compromised. On the contrary, it is increased. And with such approaches, welfare is increased as well.

To broaden the viewscreen, consider three stylized examples:

1. *Luke has heart disease, and he needs to take various medications. He wants to do so, but he is sometimes forgetful. His doctor sends him periodic text messages. As a result, he takes the medications. He is very glad to receive those messages.*
2. *Meredith has a mild weight problem. She is aware of that fact, and while she does not suffer serious issues of self-control, and does not want to stop eating the foods that she enjoys, she does seek to lose weight. Because of a new law, many restaurants in her city have clear calorie labels, informing her of the caloric content of various options. As a result, she sometimes chooses low-calorie offerings — which she would*

¹⁷ See Susan Parker, *Esther Duflo Explains Why She Believes Randomized Control Trials Are So Important*, CENT. FOR EFFECTIVE PHILANTHROPY (June 23, 2011), <http://cep.org/esther-duflo-explains-why-she-believes-randomized-controlled-trials-are-so-vital/>.

not do if she were not informed. She is losing weight. She is very glad to see those calorie labels.

3. *Rita teaches at a school, which has long offered its employees the option to sign up for a retirement plan. Rita believes that signing up would be a terrific idea, but she has not gotten around to it. She is somewhat embarrassed about that. Last year, the school, switched to an automatic enrollment plan, by which employees are defaulted into the school's plan. They are allowed to opt out, but Rita does not. She is very glad that she has been automatically enrolled in the plan.*

In all of these cases, the relevant intervention increases navigability. Nor is there is anything unfamiliar about such cases. On the contrary, they capture a great deal of the real-world terrain of nudging, both by governments and by the private sector. Choosers have a goal, or an assortment of goals, and the relevant choice architecture can make it easier or harder for them to achieve it or them. Insofar as we understand the “as judged by themselves” criterion by reference to people’s antecedent preferences, that criterion is met.

Note that it would be easy to design variations on these cases in which nudges *failed* that criterion, because they would make people worse off by their own lights. Richard Thaler uses the term “sludge” for practices that reduce navigability. Unfortunately, and sometimes tragically, sludge is everywhere. Sludge reduces freedom, in the sense that it makes it harder for people to get where they want to go. We might go further: In many nations, sludge is a principal obstacle to freedom.

We could complicate the cases of Luke, Meredith, and Rita by assuming that they have clear antecedent preferences, that the nudge is inconsistent with those preferences, but that as a result of the nudge, their preferences are changed. For example:

Jonathan likes talking on his cell phone while driving. He talks to friends on his commute to work, and he does business as well. As a result of a set of vivid warnings, he has stopped. He is glad. He cannot imagine why anyone would talk on a cell phone while driving. In his view, that is too dangerous.

After the nudge, Luke, Meredith, Rita, and Jonathan believe themselves to be better off. Cases of this kind raise the question whether the “as judged by themselves” criterion requires reference to ex ante or ex post preferences. That is an excellent question, which might be answered by making direct inquiries into people’s welfare; I will turn to that question below. My main point is that as originally given, the cases of Luke, Elizabeth, and Rita are straightforward. Such cases are common.

Self-Control

Philosophers, economists, psychologists, lawyers, and others have long been concerned with a variety of problems that go under the general rubric of “self-control.” Philosophers speak of weakness of will, or *akrasia*, which refers to susceptibility to temptation — too much food, too much drink, too much sex, too little concern for the future. Economists and psychologists speak

of “present bias” and unrealistic optimism, which may lead people to injure their long-term selves. They also refer to System 1 and System 2, where System 1 is the automatic system (focused on today and tomorrow) and System 2 is the deliberative system (capable of taking the long view).

A small story: I have an eight-year-old son, named Declan, who loves toys. Whenever we pass a toy store, he wants to stop. One day, I told him about the difference between the two systems, and I explained that while his System 1 wants toys, his System 2 is well aware that he has no need for more toys. (Doesn’t every good father explain that?) For a while, he understood the point, and it seemed to help. But after a month, he asked me, “Daddy, do I even *have* a System 2?”

Self-control problems raise conceptual, empirical, and normative challenges. One question is whether those who indulge themselves, today or this month, suffer from a self-control problem, or instead have a rational mantra: “Enjoy life now. This is not a rehearsal.” Another question is whether purported solutions to self-control problems will make the situation better rather than worse. Some cures can be worse than the disease.

Consider these haunting, ambivalent words from Byatt’s heroine, Charlotte LaMotte, writing to her dying lover, Randolph Ash: “I would rather have lived alone, so, if you would have the truth. But since that might not be — and is granted to almost none — I thank God for you — if there *must* be a Dragon — that He was You —.”¹⁸ Readers of *Genesis* have long pondered whether the choices of Adam and Eve, in the Garden of Eden, reflected a fatal inability to resist temptation (the conventional view), or something very different, such as an exercise of autonomy or an honorable desire for knowledge (and in an important sense freedom). Was the serpent only or altogether a villain? Was he a villain at all? The conventional view has triumphed in most circles, but the appeal of the alternative view helps account for the enduring power of *Genesis*.

Notwithstanding these debates, and LaMotte’s ambivalence, there is no question that many people agree, before and after the fact, that interventions can help them overcome self-control problems, even if they preserve freedom of choice. Consider these cases:

1. *Ted smokes cigarettes. He wishes that he had not started, but he has been unable to quit. His government has recently imposed a new requirement, which is that cigarette packages must be accompanied with graphic images, showing people with serious health problems, including lung cancer. Ted is deeply affected by those images; he cannot bear to see them. He quits, and he is glad.*
2. *Joan is a student at a large university. She drinks a lot of alcohol. She enjoys it, but not that much, and she is worried that her drinking is impairing her performance and her health. She says that she would like to scale back, but for reasons that she does not entirely understand, she has found it difficult to do so. Her university recently embarked on an educational campaign to reduce drinking on campus, in which it*

¹⁸ BYATT, *supra* note 2, at 546.

(accurately) notes that four out of five students drink only twice a month or less. Informed of the social norm, Joan finally resolves to cut back her drinking. She does, and she is glad.

In these cases, the chooser suffers from a self-control problem and is fully aware of that fact. Ted and Joan can be seen as both planners, with second-order preferences, and doers, with first-order preferences. A nudge helps to strengthen the hand of the planner. It is possible to raise interesting philosophical and economic questions about akrasia and planner-doer models, but insofar as Ted and Joan welcome the relevant nudges, and do so ex ante as well as ex post, the “as judged by themselves” criterion is met.

We should underline the fact that when outsiders, including choice architects, contend that choosers suffer from a self-control problem, they ought to be humble. Any outsider faces an epistemic problem. Choosers might not, in fact, be adversely affected by present bias or inertia; they might love what they are doing, even if it harms their future self, and they might be making a rational, or rational enough, tradeoff between now and later. Consider a delicious meal, a wild night, two weeks off, an apparently incautious love affair. Life is not a rehearsal, and planners need to do. The only point is that in important cases, self-control problems are serious and real, and choosers will acknowledge that fact.

In a sense, solutions to self-control problems require their own GPS devices and so can be seen to involve navigability. Appropriate interventions lead them where they want to go (at least on reflection). They promote freedom. But for choosers who face self-control problems, the underlying challenge is qualitatively distinctive — it is not in any sense a problem of insufficient knowledge — and they recognize that fact. (And recall LaMotte’s words. Her System 1 did not regret what happened, and sometimes, System 1 rules the roost.)

What Do People Think?

In the cases I am exploring, the “as judged by themselves” criterion is met. But do people know that they face a self-control problem? That is an empirical question, of course, and my own preliminary research suggests that the answer is “yes.” On Amazon’s Mechanical Turk, I asked about 200 people this question:

Many people believe that they have an issue, whether large or small, of self-control. They may eat too much, they may smoke, they may drink too much, they may not save enough money. Do you believe that you have any issue of self-control?

No less than 70 percent said that they did (55 percent said “somewhat agree,” while 15 percent said “strongly agree”). Only 22 percent disagreed. (Eight percent were neutral.)

This is a preliminary test, of course. Whatever majorities say, the cases of Ted and Mary capture a lot of the territory of human life, as reflected in the immense popularity of programs designed to help people to combat addiction to tobacco and alcohol. We should agree that nudges

that do the work of such programs, or that are used in such programs, are likely to satisfy the “as judged by themselves” criterion.

The Hardest Problems

There are much harder cases. In some of them, it is not clear if people have antecedent preferences at all. Perhaps they do not. In others — as in the case of Jonathan — their ex post preferences are an artifact of, or constructed by, the nudge. Sometimes these two factors are combined (as marketers are well aware). We are speaking here of “endogenous preferences,” and in particular to *preferences that are endogenous to the relevant choice architecture*. As Amos Tversky and Richard Thaler put it long ago, “values or preferences are commonly constructed in the process of elicitation.”¹⁹ If so, how ought the “as judged by themselves” criterion to be understood and applied? That is a singularly challenging question.

For example:

1. *George cares about the environment, but he also cares about money. He currently receives his electricity from coal; he knows that coal is not exactly good for the environment, but it is cheap, and he does not bother to switch to wind, which would be slightly more expensive. He is quite content with the current situation. Last month, his government imposed an automatic enrollment rule on electricity providers: People will receive energy from wind, and pay a slight premium, unless they choose to switch. George does not bother to switch. He says that he likes the current situation of automatic enrollment. He approves of the policy and he approves of his own enrollment.*
2. *Mary is automatically enrolled in a Bronze Health Care Plan — it is less expensive than Silver and Gold, but it is also less comprehensive in its coverage, and it has a higher deductible. Mary prefers Bronze and has no interest in switching. In a parallel world (a lot like ours, but not quite identical), Mary is automatically enrolled in a Gold Health Care Plan — it is more expensive than Silver and Bronze, but it is also more comprehensive in its coverage, and it has a lower deductible. Mary prefers Gold and has no interest in switching.*
3. *Thomas has a serious illness. The question is whether he should have an operation, which is accompanied with potential benefits and potential risks. Reading about the operation online, Thomas is not sure whether he should go ahead with it. Thomas’ doctor advises him to have the operation, emphasizing how much he has to lose if he does not. He decides to follow the advice. In a parallel world (a lot like ours, but not quite identical), Thomas’s doctor advises him not to have the operation, emphasizing how much he has to lose if he does. He decides to follow the advice.*

¹⁹ Amos Tversky & Richard H. Thaler, *Anomalies: Preference Reversals*, 4 J. ECON. PERSP. 201, 210 (1990).

A note here on the concept of parallel worlds: Science fiction writers like to speak of such worlds, showing that with a little twist or a small alteration, our lives, our cities, our nations, our entire world might be very different. (From *Possession*: “There are things that happen and leave no discernible trace, are not spoken or written of, though it would be very wrong to say that subsequent events go on indifferently, all the same, as though such things had never been.”²⁰) Parallel worlds are intriguing for many reasons, and the very idea is (I think) deep, because it highlights the omnipresence of contingency. One such reason is that we — you and I — might have been, or be, quite happy in multiple worlds, even if we are quite happy in our own. For the “as judged by themselves” criterion, that is a serious problem.

History is only run once, but nudges can be taken to create parallel worlds. In the latter two cases, Mary and Thomas appear to lack an antecedent preference; what they prefer is an artifact of the default rule (in the case of Mary) or the framing (in the case of Thomas). George’s case is less clear, because he might be taken to have an antecedent preference in favor of green energy, but we could easily understand the narrative to mean that his preference, like that of Mary and Thomas, is endogenous to the default rule.

These are the situations on which I am now focusing: *What people like is a product of the nudge*. Their preference is constructed by it. After being nudged, they will be happy and possibly grateful. One reason might be learning. Another reason might be reduction of cognitive dissonance; people might reduce dissonance in a way that makes them satisfied with the new status quo, *whatever it is*. If so, it is hard to see the “as judged by themselves” criterion as sufficient, because by hypothesis, people are satisfied only because dissatisfaction is unpleasant or unbearable, and because they would be satisfied either way.

We have also seen that even if people have an antecedent preference, the nudge might change it, so that they will be happy and possibly grateful even if they did not want to be nudged in advance. The most extreme cases involve “big decisions”²¹ or “transformative experiences,”²² in which people’s identity, and their preferences and values, are at stake; once they make certain decisions, or are nudged to make them, what they care about, and who they most deeply are, are different from what they were before. People might, for example, decide to have a romance, to get married, to have children, to change occupations, to change cities. Some of these decisions turn out to be defining. (From *Possession*: “This is where I have always been coming to. Since my time began. And when I go away from here, this will be the midpoint, to which everything ran, before, and *from* which everything will run. But now, my love, we are *now*, and those other times are running elsewhere.”²³)

In the philosophical literature, transformative experiences are often assessed from the first-person perspective: How should you, or I, choose between the status quo and undergoing some kind of transformation? The question I am posing here is different: How should we evaluate transformative nudges?

²⁰ BYATT, *supra* note 2, at 552.

²¹ See EDNA ULLMANN-MARGALIT, *NORMAL RATIONALITY* (2017).

²² See L.A. PAUL, *TRANSFORMATIVE EXPERIENCE* (2015).

²³ BYATT, *supra* note 2, at 309.

In extreme and less extreme cases, application of the “as judged by themselves” criterion is much less simple whenever people’s preferences are an artifact of nudging. Choice architects cannot contend that they are merely vindicating choosers’ ex ante preferences. If we look ex post, people do think that they are better off, and in that sense the criterion is met. For use of the “as judged by themselves” criterion, the challenge is that *however Mary and Thomas are nudged, they will agree that they are better off*. In my view, there is no escaping at least some kind of welfarist analysis in choosing between the two worlds in the cases of Mary and Thomas. We need to ask what kind of approach makes people’s lives go better.²⁴

Narrowing Options

I will get to that question in due course. Let us note preliminarily that the “as judged by themselves” criterion remains relevant in the sense that it constrains what choice architects can do, even if it does not specify a unique outcome (as it does in cases in which people have clear ex ante preferences and in which the nudge does not alter them).

Recall that it is psychologically reasonable (often) to think that choosers have antecedent preferences, but that because of a lack of information or a behavioral bias, their choices will not satisfy them. (See the cases of Luke, Meredith, and Rita.) To be sure, it is imaginable that some forms of choice architecture will affect people who do have information or who lack such biases; an error-free cafeteria visitor might grab the first item she sees, because she is busy, and because it is not worth it to her to decide which item to choose. Consider this case:

Gretchen enjoys her employer’s cafeteria. She tends to eat high-calorie meals, but she knows that, and she likes them a lot. Her employer recently redesigned the cafeteria so that salads and fruits are the most visible and accessible. She now chooses salad and fruit, and she likes them a lot.

By stipulation, Gretchen suffers from no behavioral bias, but she is affected by the nudge. But in many (standard) cases, behaviorally biased or uninformed choosers will be affected by a nudge, and less biased and informed choosers will not; a developing literature explores how to proceed in such cases, with careful reference to what seems to me a version of the “as judged by themselves” criterion.²⁵

In Gretchen’s case, and all those like it, the criterion does not leave choice architects at sea: If she did not like the salad, the criterion would be violated. From the normative standpoint, it may not be entirely comforting to say that nudges satisfy the “as judged by themselves” criterion if choice architects succeed in altering the preferences of those whom they are targeting.

²⁴ In this respect I depart from Ullmann-Margalit and Paul, who do not approach the question in this way, and who are speaking of a subclass of the cases on which I am focusing here. I cannot engage their powerful arguments in this space.

²⁵ Jacob Goldin, *Which Way to Nudge? Uncovering Preferences in the Behavioral Age*, 125 YALE L.J. 226 (2015); Jacob Goldin & Nicholas Lawson, *Defaults, Mandates, and Taxes: Policy Design with Active and Passive Decision-Makers*, 18 AM. J. L. & ECON. 438 (2016).

(Is that a road to serfdom? Recall the chilling last lines of George Orwell's 1984: "He had won the victory over himself. He loved Big Brother.") But insofar as we are concerned with subjective welfare, it is a highly relevant question whether choosers believe, ex post, that the nudge has produced an outcome of which they approve.

Two Approaches

Now let us engage the largest questions. If people would be happy, after the fact, with the outcome produced by two or more interventions, then which interventions ought people to choose? I have said that the "as judged by themselves" criterion is indeterminate. In these circumstances, we can imagine two approaches.

Follow the right choosers

The first approach would be to ask about the actual choices of people who are (a) informed and (b) unaffected by the nudge.²⁶ On this approach, choice architects would attempt to follow those choices. They would not make decisions on their own. They would track what certain people — the right choosers — actually do. We can associate this approach with John Stuart Mill, emphasizing his commitment to respecting people's choices about what would promote their welfare. Consider this possibility:

In a large population of choosers, a subset of people chooses a particular health care plan, no matter how they are nudged. They are well-informed. In terms of what they care about, they are not different, in any relevant respect, from people who are highly susceptible to nudges, and who are happy, after the fact, whether or not they are nudged.

Or consider this one:

In a large community of choosers, some shoppers choose particular kinds of bread, no matter where the bread is placed. They are well-informed; they happen to like those kinds of bread. In terms of what they care about, they are not relevantly different from other people, in their community, who are highly susceptible to nudges and who are happy, after the fact, whether or not they are nudged.

It is both tempting and plausible to think that in cases of this kind, we have a way out of our dilemma. We should follow the judgments of choosers who are at once informed and consistent. The reason is that such choosers are in an excellent position to know what is best. Inconsistent choosers, or those who are susceptible to being nudged, should take their guidance from consistent choosers, or those who are not susceptible to being nudged. On this approach, choice architects are comparatively modest. They act on the basis not of their own assessments, but the assessments of choosers whom they have reason to trust.

²⁶ See note 24 for valuable discussion; see also Jacob Goldin, *Libertarian Quasi-Paternalism*, 82 MISSOURI L. REV. 669 (2017), from which I have learned a great deal.

Welfare, directly

That approach sounds reasonable, but there is another option. It would focus directly on people's welfare, and it would not make people's choices authoritative, even if they are informed and consistent. For purposes of exposition, we can associate this approach with Jeremy Bentham. (I do not mean to suggest that Bentham would have endorsed this approach, or that the criterion for welfare is utilitarianian.) On what I am calling the Benthamite approach, the question is *which approach really does increase people's welfare, suitably defined*. This approach imposes more serious burdens on choice architects. It requires them to ask and answer the welfare question, rather than to identify and track the behavior of informed choosers.²⁷ It requires them to engage challenging normative and empirical issues. (In cases of transformative experiences, those issues are especially challenging, because we have to ask what kind of life is best.)

How should we choose between the Millian approach and the Benthamite approach? The ultimate criterion must be Benthamite, which is to say human welfare is ultimately what matters — but that does not at all mean that the Benthamite approach enjoys an easy victory.

Suppose we think that informed, consistent choosers know what they are doing. Suppose we insist that because they are informed and consistent, we really can trust them. Suppose finally that choice architects, trying to decide what approach would promote people's welfare, would make plenty of mistakes. Perhaps they do not know enough. Perhaps their motivations are not pure. On those assumptions, we would do well to follow Mill.

In some contexts, those assumptions are the right ones. But not in all contexts. Choosers might be informed and consistent, but they might suffer from a behavioral bias, such as overconfidence. Their own choices might not promote their welfare. Choice architects might be trustworthy, or at least trustworthy enough. If so, they should not follow the choices of informed, consistent choosers. By hypothesis, they know better. Some cases are relatively mundane on this count; knowledge of facts is enough. The cases of George, Mary, and Thomas are arguable examples. Some cases are far more challenging, especially if they involve large questions about people's identity and character.

Coercion

Turn now to one final topic: coercion.

Might freedom of choice fail to promote welfare? Every member of the human species knows that the answer is sometimes yes. If people suffer from unrealistic optimism, limited attention, or a problem of self-control, and if the result is a serious welfare loss, there is an argument for some kind of mandate. To see the argument, let us engage a somewhat technical question: whether to mandate fuel economy standards for automobiles. If the question seems too

²⁷ There is of course a difference between the situation of choosers, deciding whether to make decision that alters their preferences, or that is big or transformative, and the situation of choice architects, deciding whether to encourage or discourage such a choice. I am focusing on the latter here.

technical, we can substitute other questions, such as whether to ban smoking, to mandate savings, to forbid boxing, or to impose soda taxes.

The freedom-preserving approach, associated with Mill, is to provide information: Tell consumers about the economics savings. In this context, however, there is a serious risk that a nudge will be inadequate. Even with the best fuel economy label in the world, consumers might well be insufficiently attentive to those benefits at the time of purchase, not because they have made a rational judgment that they are outweighed by other factors, but simply because most of their focus is on other variables. (How many consumers think about time-savings when they are deciding whether to buy a fuel-efficient vehicle?)

If so, a suitably designed fuel economy mandate — hard paternalism, and no mere nudge — might end up producing an outcome akin to what would be produced by consumers who are at once informed and attentive. If the benefits of the mandate greatly exceed the costs, and if there is no significant consumer welfare loss (in the form, for example, of reductions in safety, performance, or aesthetics), then there is reason to believe that the mandate does serve to correct a behavioral market failure.

Of course we should be cautious before accepting that conclusion. Behavioral biases have to be demonstrated, not simply asserted; perhaps most consumers do pay a lot of attention to the benefits of fuel-efficient vehicles. The government's numbers, projecting costs and benefits, might be wrong. It is important to emphasize that consumers have diverse preferences with respect to vehicles, and regulation might end up decreasing consumers' access to vehicles with attributes that many of them prefer. With these qualifications, the argument for fuel economy standards, made by reference to behavioral market failures, is at least plausible. In this context, nudges (in the form of an improved fuel economy label) and mandates (in the form of standards) might march hand-in-hand.

This conclusion raises a final question: Should we depart from the “as judged by themselves” standard? If that question is meant to doubt whether people's judgments, *before the fact*, are always authoritative, the answer is simple: They are not! People might not welcome a mandate even though it is very much in their interests. If the question is whether people's judgments, *after the fact*, might not be authoritative, the answer is less simple. If we are concerned about people's welfare, it is surely relevant, and not at all a good sign, that people reject a mandate or a ban. In a free society, the presumption should be that they are right. But presumptions can be rebutted. If the issue involves serious harm, and if the evidence is very clear, we will have to abandon the “as judged by themselves” standard.

Reluctantly, but still.

Final Words

Countless interventions increase navigability, writ large, in the sense that they enable people to get where they want to go, and therefore enable them to satisfy their antecedent preferences. Many other interventions, helping to overcome self-control problems, are warmly

welcomed by choosers, and so are consistent with the “as judged by themselves” standard. Numerous people acknowledge that they suffer from such problems.

When people lack antecedent preferences or when those preferences are not firm, and when an intervention constructs or alters their preferences, the “as judged by themselves” standard is more difficult to operationalize, and it may not lead to a unique solution. But it restricts the universe of candidate solutions, and in that sense helps to orient choice architects. To resolve hard questions, it might make sense to see what informed, consistent choosers do, or instead to make direct inquiries into welfare.

The first approach is best unless choosers suffer from a behavioral bias — and if choice architects cannot be trusted. The second is best if choosers suffer from a behavioral bias — and if choice architects can be trusted. For the future, we need far more careful consideration of the ingredients of well-being, informed by evidence as well as by theory. We need the arts and the humanities, social science, law and theology.

Some lines from John Milton, in *Paradise Lost*, a tale of freedom, writing about Adam and Eve, who have succumbed to temptation and lost everything, and been expelled by God from the Garden of Eden²⁸:

*Some natural tears they dropped but wiped them soon;
The world was all before them, where to choose
Their place of rest, and providence their guide;
They hand in hand with wand’ring steps and slow,
Through Eden took their solitary way.*

Recall finally the passage from Byatt’s *Possession*, also a tale of freedom, of a fortunate fall, and of a uniquely human kind of joy:

In the morning, the whole world had a strange new smell. It was the smell of the aftermath, a green smell, a smell of shredded leaves and oozing resin, of crushed wood and splashed sap, a tart smell, which bore some relation to the smell of bitten apples. It was the smell of death and destruction, and it smelled fresh and lively and hopeful.

²⁸ For a brilliant discussion, see chapter 11 of STEPHEN GREENBLATT, *THE RISE AND FALL OF ADAM AND EVE* (2017).