

Duke Law Journal

VOLUME 72

MAY 2023

NUMBER 8

WELFARE NOW

CASS R. SUNSTEIN†

ABSTRACT

In evaluating interventions, policymakers should consider both their welfare effects, including their effects on people’s emotional states, and their effects on distributive justice, including their effects on those at the bottom of the economic ladder. The arguments for investigating welfare effects, and effects on distributive justice, are meant as objections to efforts to evaluate behaviorally informed interventions solely in terms of (for example) revealed preferences and effects on participation rates. The arguments are also meant as a plea for investigation and specification of the effects of such interventions on experienced well-being. If interventions give people a sense of security and safety, that is a strong point in their favor; if they make people feel frightened and sad, that is a strong point against them. A central concern is that policymakers sometimes neglect the emotional impact, whether negative or positive, of behaviorally informed interventions. Personalized approaches can promote distributive goals and also target interventions to those who are most likely to be helped by them.

TABLE OF CONTENTS

I. What Matters.....	1644
II. Behaviorally Informed Interventions: Nudges and Beyond	
Nudges	1653

Copyright © 2023 Cass R. Sunstein.

† Robert Walmsley University Professor, Harvard University. This essay was the basis for a keynote address that was presented at Duke Law School in March 2023; it was also presented at the University of Cambridge at the inaugural event of the El-Erian Institute for Behavioural Economics and Policy, Cambridge, England, held in June 2022. I am grateful to Eric Xu for superb research assistance and excellent comments. I am also grateful to audiences at Duke and Cambridge for valuable comments and suggestions.

A. The Behavioral Toolbox: Nudges, Taxes, Subsidies, Mandates, Bans	1654
B. What Is Welfare?	1658
III. Benevolent Spectators	1660
A. What Is the Currency?.....	1661
B. Prioritarianism.....	1663
IV. Welfare and Distribution.....	1665
V. Targeting and Personalization	1670
Conclusion.....	1672

In time the world will cool and everything will die; but that is a long time off still, and its present value at compound discount is almost nothing. Nor is the present less valuable because the future will be blank. Humanity, which fills the foreground of my picture, I find interesting and on the whole admirable. I find, just now at least, the world a pleasant and exciting place. You may find it depressing; I am sorry for you, and you despise me. But I have reason and you have none; you would only have a reason for despising me if your feeling corresponded to the fact in a way mine didn't. But neither can correspond to the fact. The fact is not in itself good or bad; it is just that it thrills me but depresses you. On the other hand, I pity you with reason because it is pleasanter to be thrilled than to be depressed, and not merely pleasanter but better for all one's activities.

Frank Ramsey¹

I. WHAT MATTERS

I will be making one general argument here. I hope that it will not be controversial. It will serve as an umbrella for three more particular arguments, and they might be controversial.

The general argument is that interventions informed by behavioral science,² like all other interventions, should be evaluated by reference to their effects on human welfare. The first of the particular

1. FRANK PLUMPTON RAMSEY, FOUNDATIONS OF MATHEMATICS 291–92 (R.B. Braithwaite ed., 1931).

2. See generally CASS R. SUNSTEIN, BEHAVIORAL SCIENCE AND PUBLIC POLICY 1–2 (Robin Boadway, Frank A. Cowell & Massimo Florio eds., 2020) (defining “behavioral insights” as “behavioral findings” from “cognitive psychology, social psychology, and behavioral economics,” and examining how behavioral insights could inform public policy).

arguments, and the one I mean to press most heavily,³ is that it is important to focus on the effects of interventions on people's *emotional states*. If a behaviorally informed intervention makes people sad or scared, that is a point against it. Of course, it is also true that such an intervention might have significant benefits—for example, by lengthening lives—and the benefits might greatly outweigh the costs. But the effects of interventions on emotional states are part of the picture, and they should not be neglected.

The second of my particular arguments is that behaviorally informed interventions, like all other interventions, should be evaluated in terms of distributive justice, and, in particular, in terms of their effects on those at the bottom of the welfare ladder. Some interventions have beneficial aggregate effects, but they do not help, and might hurt, those at the bottom. That is a serious problem.

The third of the particular arguments is that personalized or targeted interventions have a great deal of promise in increasing human welfare. One reason is that they can increase the benefits of interventions in terms of people's emotional states; interventions can be applied to those who like them and not to those who hate them. Another reason is that personalized interventions can promote distributive justice. An intervention might, for example, be applied to those who are least likely to lose from it (including by being distressed or anguished), and it might also be applied, or not applied, to those at the bottom of the welfare ladder, depending on whether they would be helped or hurt.

To see how these arguments might work, consider three examples.

Begin with the case of “home energy reports,” by which people receive information about their energy consumption and how their usage compares to that of their neighbors. Such reports have a range of effects. They help people save money; they produce reductions in air pollution (including in greenhouse gases); they also cost money to disseminate. If we subtract the material costs from the material benefits, the net welfare benefit of home energy reports is substantial.⁴

3. I do so not because it is the most important but because it is the most neglected. *See, e.g.,* John Bronsteen, Christopher Buccafusco & Jonathan S. Masur, *Well-Being Analysis vs. Cost-Benefit Analysis*, 62 DUKE L.J. 1603, 1603 (2013) (proposing a well-being analysis as an alternative to the primary public policy tool of cost-benefit analysis, which “convert[s] preferences . . . into dollar figures”).

4. *See* Hunt Allcott & Judd B. Kessler, *The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons*, 11 AM. ECON. J.: APPLIED ECON. 236, 238 (2019) (estimating

But a number of people do not *like* receiving home energy reports.⁵ (Would you like to receive one?) When people do not like receiving such reports, it might be because they do not like having to try to understand them. It might be because they do not enjoy the process of thinking about whether and how to save energy, or of taking steps to save energy. According to a prominent study, traditional approaches to evaluation, focusing only on material impacts (the strictly economic costs and benefits) and neglecting emotional impacts, overstate the welfare gains of home energy reports by a factor of 3.7, which means that the overall welfare gains might be overstated by as much as \$620 million.⁶ The central reason is that the emotional impacts are often negative; many people do not like receiving the reports.

Turn now to the case of calorie labels. At first glance, the purely material effects of those labels are straightforward. They might lead to healthier food choices, which is, of course, a benefit; their production imposes costs. According to a prominent study, the benefits of such labels do indeed exceed the costs.⁷ But there is an important wrinkle: many people are saddened by the labels (they contain bad news!). For that reason, the labels impose a kind of “emotional tax” on people with low levels of self-control with respect to eating. Such people receive fewer benefits from the intervention and perhaps no benefits at all; they might be sadder and perhaps ashamed, but they eat exactly what they would have eaten without the labels.⁸ If so, the people who most need calorie labels in terms of health, and whom in that sense the labels are intended to benefit, might be net losers; they might not change their behavior, and they might feel worse and more agitated as a result of the labels.

Turn finally to the case of saving for retirement. Why do people not save more? What would help? Empirical work finds a host of contributing factors.⁹ The most intriguing involves a combination of (1)

that a home energy reports program “increase[d] social welfare by \$0.77 per household” in its second year).

5. See *id.* (noting 34 percent of people did not want to receive a report even if it were free).

6. *Id.* at 238–39.

7. Linda Thunström, *Welfare Effects of Nudges: The Emotional Tax of Calorie Menu Labeling*, 14 JUDGMENT & DECISION MAKING 11, 11 (2019) (noting that, overall, such labels “positively affect[] consumer welfare”).

8. *Id.* at 17, 20.

9. See, e.g., Saurabh Bhargava & Lynn Conell-Price, *Serenity Now, Save Later? Evidence on Retirement Savings Puzzles from a 401(k) Field Experiment* (Mar. 2022) (unpublished manuscript), https://ssrn.com/abstract_id=4056407 [<https://perma.cc/XJ2Q-Q3M2>] (providing evidence

anxiety about the present and (2) optimism about the future, which leads people to be simultaneously nervous about losing current income and optimistic about their own economic futures. It follows that mandatory saving, or even a nudge in favor of saving, might produce both fear and distress.¹⁰

In numerous cases, behaviorally informed interventions may or may not have desirable distributional effects, and they may or may not make people feel sad or distressed. To see the sheer scope of the problem, consider the following cases:

(1) Concerned about the problem of obesity, a government imposes a tax on sugar-sweetened beverages.¹¹ Poor people do not want to pay the tax, and in large numbers they cease buying sugar-sweetened beverages. They might suffer serious welfare losses as a result.

(2) To reduce greenhouse gas emissions and to save consumers money, a government imposes aggressive energy efficiency requirements on appliances.¹² These requirements increase the selling prices of the products to which they are applied. A given amount of money is worth more to the poor than to the wealthy. Poor people are especially reluctant to buy those more expensive products, which means that the energy efficiency requirements operate as a kind of regressive tax.

for four explanations: low financial and retirement literacy, “confusion about plan details,” employers’ complex enrollment processes, and privileging the present over the future).

10. Regulators could respond by addressing the relevant anxiety, perhaps with [d]ual-account proposals that supplement the illiquidity of a traditional 401(k) with a more-liquid account designed to address near-term financial concerns. Such a plan could direct employee contributions into a highly-liquid account that provides emergency liquidity before automatically transferring accumulated funds, above a certain threshold, into a less-liquid account long-term account.

Id. at 47–48.

11. See Hunt Allcott, Benjamin B. Lockwood & Dmitry Taubinsky, *Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence*, 33 J. ECON. PERSPS. 202, 202, 219 (2019) [hereinafter Allcott et al., *Should We Tax Sugar-Sweetened Beverages?*] (evaluating taxes on sugar-sweetened beverages and noting that, in light of the taxes, “low-income households reduce[d] sugar-sweetened beverage consumption”).

12. See Don Fullerton & Chi L. Ta, *Costs of Energy Efficiency Mandates Can Reverse the Sign of Rebound*, J. PUB. ECON., Aug. 2020, at 1, 1 (analyzing rebound effects from increased energy efficiency mandates).

(3) An employer automatically enrolls employees in a savings program with a 9 percent contribution rate.¹³ Many low-income employees find that their take-home pay has been reduced. They do not understand exactly why, but they are significantly less well off. After a year, some (but not all) respond by opting out.

(4) An employer allows employees to compete for certain promotions. Noticing that men are more likely than women to enter such competitions, the employer adopts a new policy: employees are automatically entered into those competitions, with the right to opt out. The effect of the policy is to eliminate the previous disparity between men and women.¹⁴

I have said that to evaluate interventions of these kinds, the central question is whether an intervention improves human welfare.¹⁵ An intervention that helps people live longer lives has a strong point in its favor. The same can be said if an intervention reduces human distress. Of course, it is true that longevity is not the only thing that matters to human welfare. Of course, it is also true that while reducing distress is a good thing, it is not the *only* good thing. Longevity contributes to human welfare, and human welfare increases if people experience less in the way of distress. But welfare itself is what matters, and welfare emphatically includes people's emotional states, which is why emotional states matter.¹⁶ If people feel frightened, anxious, or miserable, they are to that extent less well off; fear, anxiety, and misery are costs, and those costs should be reduced, not increased.

The idea of welfare is a broad and capacious one, and it can be specified in many different ways.¹⁷ For purposes of law and regulation, we can make a great deal of progress without fussing so much over the

13. See Shlomo Benartzi & Richard H. Thaler, *Behavioral Economics and the Retirement Savings Crisis*, 339 SCIENCE 1152, 1152 (2013) (identifying automatic enrollment as a key feature of promoting adequate retirement savings).

14. See generally Joyce C. He, Sonia K. Kang & Nicola Lacetera, *Opt-Out Choice Framing Attenuates Gender Differences in the Decision To Compete in the Laboratory and in the Field*, PNAS, Oct. 11, 2021, at 1 (discussing multiple studies to this effect).

15. I am bracketing questions about autonomy, though of course they might be important in many cases.

16. I am aware that this proposition is inconsistent with widespread intuitions. On the role of those intuitions, see generally Cass R. Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCIS. 531 (2005) (urging that many moral intuitions can be seen as heuristics for what matters).

17. See Amartya Sen, *Utilitarianism and Welfarism*, 76 J. PHIL. 463, 471–72 (1979) (offering a capacious conception of welfare, broader than utilitarianism).

appropriate specification. Much of the time, we can reach incompletely theorized agreements on what is, or is not, welfare promoting. Incompletely theorized agreements reflect a consensus on what to do in the absence of a consensus on the underlying theory that justifies doing it.¹⁸

Preference satisfaction matters to welfare, but preference satisfaction is not the only thing that matters; it should not be the foundation of public policy.¹⁹ Satisfying people's preferences might make their lives worse. People might prefer something that makes them sad, sick, or dead. People might prefer Option A over Option B, but if Option B produces more welfare than Option A, we should not celebrate a situation in which everyone ends up with Option A.²⁰ The appropriate role of regulators is a separate question; if we think that regulators lack relevant information or are subject to biases of their own, we might want them to proceed with caution in second-guessing people's preferences. Perhaps regulators should not second-guess those preferences at all. But if regulators really do know that Option B makes human lives better than does Option A, they might be permitted or required to intervene in order to encourage, promote, or mandate Option B.

Overall welfare is not all that matters; the distribution of welfare matters as well.²¹ Suppose that we have a choice between two outcomes for a two-person society: (1) one person has a sensational life and one person has a horrible life; (2) both people have good lives. We might choose (2) even if the aggregate level of welfare is higher under (1). Whether we should do so depends on the appropriate theory of justice. Now suppose that we have a choice between two outcomes for a four-hundred-person society: (1) three hundred people have wonderful

18. See generally Cass R. Sunstein, *Incompletely Theorized Agreements*, 108 HARV. L. REV. 1733 (1994) (discussing pervasiveness and virtues of such agreements).

19. See generally Matthew Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FLA. ST. U. L. REV. 241 (2000) [hereinafter Adler, *Beyond Efficiency and Procedure*] (offering a broad conception of welfare and objecting to economic efficiency as a regulatory goal).

20. See generally SARAH CONLY, *AGAINST AUTONOMY: JUSTIFYING COERCIVE PATERNALISM* (2012) (arguing in favor of paternalism, even in the form of mandates).

21. See generally MATTHEW ADLER, *MEASURING SOCIAL WELFARE: AN INTRODUCTION* (2019) [hereinafter ADLER, *MEASURING SOCIAL WELFARE*] (discussing how to measure social welfare); MATTHEW ADLER, *WELL-BEING AND FAIR DISTRIBUTION: BEYOND COST-BENEFIT ANALYSIS* (2011) (exploring the limit of cost-benefit analysis in light of general commitment to well-being).

lives, fifty people have good lives, and fifty people have horrible lives; (2) two hundred people have wonderful lives and two hundred people have good lives. We might choose (2) even if the aggregate level of welfare is higher under (1). Whether we should do so depends on the appropriate theory of justice. I mean to insist here on the centrality of both welfare and its distribution, but I do not mean to specify any theory of justice.

Some regulatory interventions are progressive; they have especially beneficial effects on those at the bottom of the economic ladder. Other interventions are regressive; they have especially harmful effects on those at the bottom of the economic ladder. Still others help those at the top but not those at bottom, thus increasing welfare while decreasing equality.²² If we understand “progressive” and “regressive” by reference to effects on other demographic groups (people with disabilities, the elderly, people of color, women), we can easily find diverse kinds of progressive interventions²³ and regressive interventions.²⁴

There is a view, of course, that regulators should focus on welfare maximization, and that distributive matters are for the tax system.²⁵ This argument has been contested on multiple grounds, and I do not accept it here.²⁶ With respect to behaviorally informed interventions, it

22. See Nicole Wetsman, *Health Apps and Wearables Help Rich People the Most, Study Finds*, VERGE (Nov. 16, 2021, 12:54 PM), <https://www.theverge.com/2021/11/16/22785163/digital-health-physical-activity-inequity> [<https://perma.cc/8ZZ7-2DDR>] (discussing a study showing “that apps and wearables only helped bump physical activity levels for people with high socioeconomic status”).

23. See Kellen Mrkva, Nathaniel A. Posner, Crystal Reeck & Eric J. Johnson, *Do Nudges Reduce Disparities? Choice Architecture Compensates for Low Consumer Knowledge*, 85 J. MKTG. 67, 82 (2021) [hereinafter Mrkva et al., *Do Nudges Reduce Disparities?*] (concluding that choice architecture can benefit consumers overall but also reduce inequities).

24. A soda tax would be regressive along important dimensions, simply because poor people have less money. Hunt Allcott, Benjamin B. Lockwood & Dmitry Taubinsky, *Regressive Sin Taxes, With an Application to the Optimal Soda Tax*, 134 Q.J. ECON. 1557, 1559 & n. 2 (2019) [hereinafter Allcott et al., *Regressive Sin Taxes*].

25. A version of the argument can be found in Louis Kaplow & Steven Shavell, *Why the Legal System Is Less Efficient Than the Income Tax in Redistributing Wealth*, 23 J. LEGAL STUD. 667 (1994) (arguing in favor of redistribution through taxes rather than through legal system or regulation).

26. See, e.g., John D. Graham, *Regulatory Reform, Cost-Benefit Analysis, and the Poor*, REGUL. REV. (Mar. 8, 2022), <https://www.theregreview.org/2022/03/08/graham-reform-benefit-cost-analysis> [<https://perma.cc/MK6E-7Z7H>] (arguing that regulatory analysis should consider effects on poor people); Zachary Liscow, *Reducing Inequality on the Cheap: When Legal Rule Design Should Incorporate Equity as Well as Efficiency*, 123 YALE L.J. 2478 (2014) (arguing for attention to equality and equity in regulatory policy).

is clear that some choice-preserving approaches (such as default rules) will give particular help to people at the bottom of the economic ladder, and that some approaches will actually harm them. That matters. It is also true that mandates, bans, and taxes will have different effects on different groups. For example, taxes on cigarettes, alcoholic drinks, sugar-sweetened beverages, and gasoline might look regressive because they come down especially hard on poor people; at the same time, they might turn out to be progressive, in the sense that they might particularly benefit poor people. That matters as well, whether we are speaking of welfare or of distributive justice.

From what I have said thus far, it follows that to evaluate behaviorally informed interventions, it is essential to consider both their welfare effects and their effects on distributive justice. Taken in the abstract, the claim is not exactly provocative (I hope), but among other things, it helps show that impressive aggregate effects along an admittedly important margin (for example, a net increase in savings or a net decrease in obesity) tell us much less than we need to know. A pervasive concern is that behaviorally informed interventions *might have negative welfare effects on subjective well-being that are easily ignored*—as, for example, when information disclosure makes people sad or scared, or when a shift to healthier eating makes people enjoy their meals less.²⁷ At the same time, such interventions *might have positive effects on subjective well-being that are easily ignored*—as, for example, when information disclosure makes people feel confident and safe, or when a shift to healthier eating makes people enjoy their meals more.

An additional concern is that behaviorally informed interventions might turn out to be more appealing, or less appealing, once we disaggregate their effects on different social groups. It might be, for example, that an anti-obesity policy has, on net, particularly high welfare benefits for poor people. It might also be, for example, that energy efficiency requirements have, on net, negative welfare effects for poor people. Some behaviorally informed interventions might impose real harms on people of color, or they might be especially beneficial to them.

27. See Cass R. Sunstein, *Ruining Popcorn? The Welfare Effects of Information*, 58 J. RISK & UNCERTAINTY 121, 123 (2019) (describing how consumers may purchase less enjoyable products when given more information about the product).

With these concerns in mind, policymakers²⁸ should ask five specific questions when deciding whether and how to proceed with behaviorally informed interventions:

(1) *What are the aggregate effects on social welfare?* For purposes of evaluation, it is tempting to focus on increases in participation rates or on cost-effectiveness. It is also tempting to focus on “expressive values”: the kinds of statements that regulatory initiatives make. The welfare question is much more important, though it might also raise hard normative, conceptual, and empirical questions.²⁹ Cost-benefit analysis is best understood as a way of answering the welfare question, but it has serious flaws, in part because experienced well-being (defined as the level of well-being that people have when they are experiencing their lives) greatly matters (though it is not all that matters), and cost-benefit analysis is only a proxy for it.³⁰ People’s emotional states are central to an evaluation of welfare effects.³¹

(2) *What are the effects of the intervention on people’s emotional states?* This question is a subcategory of the first question, but it deserves separate emphasis. We need to know whether interventions make people feel sadder or more frightened, or safer and more secure.

(3) *Who is likely to be helped and who is likely to be hurt?* This is a plea for a distributional analysis of the effects of behaviorally informed interventions. In some cases, the analysis will be exceptionally difficult to undertake; feasibility is a constraint. But more information is generally better than less.

(4) *What are the expected effects on the least well off?* It is important to ask whether the relevant interventions help or hurt those who have the least, defined in terms of welfare.³² There is a close connection between this question and “prioritarianism,” taken up below.

28. I am assuming, of course, that policymakers are interested in good policy, not their political self-interest.

29. See W. Kip Viscusi, *Efficiency Criteria for Norms and Nudges* 24 (Pub. Choice, Working Paper No. 19-14, 2019) (discussing the questions involved in a cost-benefit analysis of “nudge” policies).

30. See generally CASS R. SUNSTEIN, *THE COST-BENEFIT REVOLUTION* (2018) (supporting cost-benefit analysis but urging that it is only a proxy for welfare).

31. See *infra* Part II.B.

32. See Graham, *supra* note 26.

(5) *Do the benefits to those helped exceed the costs to those hurt?* If the gainers gain more than the losers lose, we have a point in favor of the intervention (the Kaldor-Hicks criterion,³³ for which I mean to offer, perhaps improbably,³⁴ two and a half cheers here). But the point might not be decisive if, for example, the gainers are well off to begin with, and the losers are not.

In some cases, answers to the five questions will actually strengthen the argument for behaviorally informed interventions; in other cases, such answers will weaken that argument; and in still other cases, such answers will raise a set of new issues, both normative and empirical. The most important implication is the large value of shifting toward more targeted or personalized interventions, which can often produce higher welfare benefits and prove far better on distributional grounds than mass approaches.³⁵ Targeted or personalized interventions should, I suggest, be the wave of the future for behaviorally informed policymaking. In fact, one of my main points lies there.

A note before we embark: My focus is on behaviorally informed interventions, but the same conclusions apply to interventions of all kinds. If the government is simply trying to alter incentives—for example, with subsidies—the five questions above should also be asked. They should also be asked if the government is imposing mandates or bans that have nothing at all to do with behavioral economics. I have emphasized that there are hard questions about the appropriate definition of welfare; I will leave those questions largely unanswered here, but on occasion I will identify them.

II. BEHAVIORALLY INFORMED INTERVENTIONS: NUDGES AND BEYOND NUDGES

In recent years, both public and private institutions have shown a great deal of interest in “nudges,” understood as interventions that preserve freedom of choice but that also steer people in particular

33. For discussion, see generally *supra* note 21 and accompanying text; MATTHEW D. ADLER & ERIC A. POSNER, *NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS* (2006) (not embracing the Kaldor-Hicks criterion but arguing that cost-benefit analysis is a good proxy for welfare).

34. For a series of objections, see generally Adler, *Beyond Efficiency and Procedure*, *supra* note 19 (criticizing the Kaldor-Hicks criterion on distributional and other grounds).

35. In the same vein, see Allcott & Kessler, *supra* note 4, at 239.

directions.³⁶ Nudges fall in two general categories: the educative and the architectural. Educative nudges include warnings, reminders, and the disclosure of information (such as calorie labels, allergy warnings, text reminders that bills are due, and fuel economy labels). Architectural nudges include automatic enrollment, mandatory choice, and simplification so as to highlight, and draw attention to, certain options (such as through the design of websites, forms, or in-person stores). Whether educative or architectural, nudges have often been found to have significant effects on outcomes and to be highly cost-effective.³⁷ For example, a shift from an opt-in to an opt-out design can increase participation rates by twenty-five percentage points or more.³⁸ In this Part, I will explore the behavioral toolbox (which extends well beyond nudges), the centrality of welfare and some of the conceptual and normative challenges raised by that idea, and the need to attend to the effect of interventions on people's subjective experiences.

A. *The Behavioral Toolbox: Nudges, Taxes, Subsidies, Mandates, Bans*

With respect to choice-preserving approaches, there is a large dispute over whether educative nudges should be preferred to architectural nudges, or vice versa.³⁹ Some people believe that educative nudges are better insofar as they promote people's capacity for agency.⁴⁰ "Boosting," sometimes understood as a distinctive behavioral alternative, might be taken to include educative nudges; its basic goal is to enable people to make better choices for themselves.⁴¹ Architectural nudges, by contrast, may or may not educate.

36. See generally RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: THE FINAL EDITION* (2021) (exploring updated conceptions of how nudges and choice architecture can help people make better decisions).

37. See Benartzi & Thaler, *supra* note 13, at 1152–53.

38. Jon M. Jachimowicz, Shannon Duncan, Elke U. Weber & Eric J. Johnson, *When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects*, 3 BEHAV. PUB. POL'Y 159, 165 (2019).

39. See Shai Davidai & Eldar Shafir, *Are 'Nudges' Getting a Fair Shot? Joint Versus Separate Evaluation*, 4 BEHAV. PUB. POL'Y 273, 276–77 (2020).

40. Cass R. Sunstein, *People Prefer System 2 Nudges (Kind Of)*, 66 DUKE L.J. 121, 123 (2016) ("[I]t is tempting to prefer [educative] nudges, on the assumption that they . . . promote individual agency.").

41. Ralph Hertwig & Till Grüne-Yanoff, *Nudging and Boosting: Steering or Empowering Good Decisions*, 12 PERSPS. ON PSYCH. SCI. 973, 973 (2017) ("The objective of boosts is to foster people's competence to make their own choices.").

Importantly, there has also been increasing interest in *behaviorally informed interventions that do not count as nudges* and that might take the form of subsidies, taxes, mandates, and bans.⁴² Suppose, for example, that people show “present bias” (understood as a focus on the short term and a disregard of the long term) and do not purchase energy-efficient appliances, even though doing so is in their interest. Focusing on short-term costs, people might neglect the fact that energy-efficient appliances can save them money in the long term.⁴³ An energy-efficient mandate might reduce “internalities.”⁴⁴ Or suppose that people buy high-calorie, high-sugar soft drinks because they taste good in the short term, even though consumption of such drinks imposes health risks in the long term. Some kind of internality-correcting tax might seem to be a corrective.⁴⁵ So too a subsidy of some

42. See generally CONLY, *supra* note 20 (asserting that, in some circumstances, individuals will be better off if the government prevents them from acting as they wish).

43. See, for example, this explanation from the Department of Energy:

There are several market failures or barriers that affect energy decisions generally. Some of those that affect the commercial sector specifically are detailed below. However, more generally, there are several behavioral factors that can influence the purchasing decisions of complicated multi-attribute products, such as boilers. For example, consumers (or decision makers in an organization) are highly influenced by choice architecture, defined as the framing of the decision, the surrounding circumstances of the purchase, the alternatives available, and how they're presented for any given choice scenario. The same consumer or decision maker may make different choices depending on the characteristics of the decision context (*e.g.*, the timing of the purchase, competing demands for funds), which have nothing to do with the characteristics of the alternatives themselves or their prices. Consumers or decision makers also face a variety of other behavioral phenomena including loss aversion, sensitivity to information salience, and other forms of bounded rationality. . . . These characteristics describe almost all purchasing situations of appliances and equipment, including CPBs. The installation of a new or replacement CPB in a commercial building is a complex, technical decision involving many actors and is done very infrequently, as evidenced by the CPB mean lifetime of nearly 25 years. Additionally, it would take at least one full heating season for any impacts on operating costs to be fully apparent. Further, if the purchaser of the CPB is not the entity paying the energy costs (*e.g.*, a building owner and tenant), there may be little to no feedback on the purchase.

Energy Conservation Standards for Commercial Packaged Boilers, 87 Fed. Reg. 23421, 23423 (Apr. 20, 2022) (to be codified at 10 C.F.R. pt. 431) (footnotes omitted).

44. See Hunt Allcott & Cass R. Sunstein, *Regulating Internalities*, 34 J. POL'Y ANALYSIS & MGMT. 698, 698 (2015) (defining internalities as “costs we impose on ourselves by taking actions that are not in our own best interest”); Hunt Allcott, Sendhil Mullainathan & Dmitry Taubinsky, *Energy Policy with Externalities and Internalities*, 112 J. PUB. ECON. 72, 72 (2014) (explaining “undervaluation could cause consumers to impose ‘internalities’ on themselves as they buy goods that are less energy efficient than they would choose in their private optima”).

45. See generally Allcott et al., *Regressive Sin Taxes*, *supra* note 24 (proposing an optimal commodity tax formula with a focus on internalizing both externalities and internalities).

kind, designed to encourage people to purchase electric vehicles or healthier drinks, might counteract limited attention and present bias.⁴⁶

Behaviorally informed interventions often have large effects.⁴⁷ Some nudges significantly increase participation rates.⁴⁸ Moreover, there is evidence that they can be highly cost-effective, in large part because they tend to be inexpensive.⁴⁹ An educative nudge, or a change in architecture, may or may not have a larger impact, per unit of dollar spent, than a subsidy, a tax, or a mandate.⁵⁰ It is also true that whether the focus is on externalities or internalities, a corrective tax might turn out to have a significant impact on consumption choices.⁵¹

It is increasingly clear, however, that such aggregate effects will not tell public and private institutions what they need to know. For example, an increase in participation rates does not tell us whether and to what extent an intervention is increasing social welfare (including,

46. See Ryan Bubb & Richard H. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. 1593, 1675 (2014) (noting a well-calibrated subsidy “ensures that a consumer buys the right car”).

47. See *id.* at 1674–75 (explaining behavioral insights within a market-based approach may allow policymakers to implement a fuel-efficient economy while still protecting incentives for fuel consumption).

48. See Jachimowicz et al., *supra* note 38, at 160 (noting one study that documented how “employees are 50% more likely to participate in a retirement savings program when enrollment is the default”); Peter Bergman & Todd Rogers, *The Impact of Defaults on Technology Adoption, and Its Underappreciation by Policymakers* 5 (Harv. Kennedy Sch., Working Paper No. RWP17-021, 2017), https://ssrn.com/abstract_id=3098299 [<https://perma.cc/R39H-6MAC>] (finding a massive increase in participation rates from a shift from opt-in to opt-out).

49. See Shlomo Benartzi, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon & Steven Galing, *Should Governments Invest More in Nudging?*, 28 PSYCH. SCI. 1041, 1042 (2017).

50. See Raj Chetty, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen & Tore Olsen, *Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark*, 129 Q.J. ECON. 1141, 1215 (2014) (finding that a nudge, in the form of automatic enrollment, has a much larger effect in promoting retirement savings than a significant subsidy). See generally John A. List, Matthias Rodemeier, Sutanuka Roy & Gregory Sun, *Judging Nudging: Toward an Understanding of the Welfare Effects of Nudges Versus Taxes*, FIELD EXPERIMENTS (Dec. 2022), <https://ideas.repec.org/p/feb/framed/00765.html> [<https://perma.cc/MCU7-EJ7S>] (exploring the circumstances in which nudges or taxes will have higher welfare benefits).

51. See Jonathan H. Gruber & Sendhil Mullainathan, *Do Cigarette Taxes Make Smokers Happier*, 5 ADVS. ECON. ANALYSIS & POL’Y 1, 2 (2005) (noting that “those with a propensity to smoke are significantly happier when excise taxes rise”); see also Allcott et al., *Should We Tax Sugar-Sweetened Beverages?*, *supra* note 11, at 211 (“In other words, externality benefits from a sugary drink tax are theoretically likely to be progressive, even if the financial costs are regressive.”).

but not limited to, experienced well-being⁵²), whether it is helping or hurting identifiable groups, whether it is causing serious, unintended harm to some or many, or whether it should be redesigned in some important way. If people are automatically enrolled in some program that costs a great deal and that offers little or nothing, a high participation rate is nothing to celebrate. If automatic enrollment makes people's days and weeks worse, it might be a bad idea. A corrective tax might be quite effective, but we might not know whether its overall welfare effects are beneficial, or whether it will hurt some groups while helping others.⁵³ People might shift from one kind of good to another, with the shift increasing their savings or improving their health, but the negative effects on their subjective well-being might be significant.⁵⁴ (Consider a tax on chocolate, and perish the thought.)

There are similar limits to a demonstration of cost-effectiveness.⁵⁵ Suppose, for example, that one intervention saves fifty lives at a cost of \$1, while another saves one hundred lives at a cost of \$100 million. The first is far more cost-effective, but the second is much better on welfare grounds. The magnitude of net benefits is what matters, not the cost-benefit ratio. (Saving fifty additional lives at a cost of \$99.999 million is a terrific bargain, at least if we value a statistical life at \$11.6 million, as

52. See generally PAUL DOLAN, *HAPPINESS BY DESIGN* (2014) (emphasizing the importance of experienced well-being); Daniel Kahneman, Peter P. Wakker & Rakesh Sarin, *Back to Bentham? Explorations of Experienced Utility*, 112 Q.J. ECON 375 (1997) (noting the usefulness of experienced utility in helping to “identify situations in which the assumption of consumer rationality should be applied with caution, subjected to empirical test, or avoided altogether”); David A. Schkade & Daniel Kahneman, *Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction*, 9 PSYCH. SCI. 340 (1996) (exploring how focusing on a single factor, among many, will inevitably exaggerate its impact).

53. Rahi Abouk, Charles J. Courtemanche, Dhaval M. Dave, Bo Feng, Abigail S. Friedman, Johanna Catherine Maclean, Michael F. Pesko, Joseph J. Sabia & Samuel Safford, *Intended and Unintended Effects of E-Cigarette Taxes on Youth Tobacco Use 3–4* (Nat'l Bureau of Econ. Rsch., Working Paper No. 29216, 2021), <https://www.nber.org/papers/w29216> [<https://perma.cc/6GYB-NS2U>].

54. On some of the limits of subjective well-being, see Daniel J. Benjamin, Ori Heffetz, Miles S. Kimball & Alex Rees-Jones, *What Do You Think Would Make You Happier? What Do You Think You Would Choose?*, 102 AM. ECON. REV. 2083, 2107 (2012); W. Kip Viscusi, *The Benefits of Mortality Risk Reduction: Happiness Surveys vs. The Value of a Statistical Life*, 62 DUKE L.J. 1735, 1736–37 (2013); B. Douglas Bernheim, *The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics*, 7 J. BENEFIT-COST ANALYSIS 12, 24–31 (2016).

55. See Benartzi et al., *supra* note 49, at 1044.

some parts of the U.S. government currently do.⁵⁶) To see the intuition: An investment of \$1 for a return of \$5 has a 5 to 1 ratio of benefits to costs, whereas an investment of \$1 million for a return of \$1.5 million has a 1.5 to 1 ratio of benefits to costs; but the second investment is far better because it delivers much higher net benefits, even though the ratio is not as good. Similarly, a cost-effective approach to producing an undesirable end would be nothing to celebrate; we could imagine a behaviorally informed (and cost-effective) approach to promoting gambling, smoking, or alcoholism.⁵⁷ What matters is welfare, not cost-effectiveness.

B. *What Is Welfare?*

Let us now turn to the idea of “welfare,” and how to understand that much-disputed term.

The willingness-to-pay (“WTP”) criterion, or the maximum price a consumer is willing to pay for a product or service, has been much maligned, but it has an important advantage—it recognizes that people care about a diverse assortment of goods. Choosers might care about improved safety, but that is not all they care about. If they are willing to pay X dollars and no more to reduce a mortality risk of one in one hundred thousand, we might want to defer to their judgment on the ground that choosers have an epistemic advantage over outsiders in deciding what might make their lives go well.⁵⁸ The problem, of course, is that people might choose poorly, perhaps because of an absence of information, perhaps because of a behavioral bias (such as present bias or unrealistic optimism), and their choices might not promote their welfare.⁵⁹ Then what? For conscientious regulators, health and safety matter. If people can live longer lives, have less illness and pain, and be free from serious mental health problems, they are better off.

56. John Putnam, *Memo: Guidance on the Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses – 2021 Update*, U.S. DEP’T OF TRANSP. (Apr. 4, 2021).

57. See generally Jamie Luguri & Lior Jacob Stahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43 (2021) (discussing the power of “dark patterns,” which are user interfaces whose designers knowingly confuse users, make it difficult to express actual preferences, or manipulate them into taking certain actions).

58. See Cass R. Sunstein, *Behavioral Welfare Economics*, 11 J. BENEFIT-COST ANALYSIS 196, 201–02 (2020); CASS R. SUNSTEIN, *WHY NUDGE?* 3–4 (2014).

59. See generally CONLY, *supra* note 20 (arguing for paternalistic coercion to overcome irrational decision-making behavior).

Resources matter as well; broadly speaking, richer is better than poorer.⁶⁰ I have said that subjective well-being matters; if people are scared or sad, their experiences are worse, and other things equal, it is generally better that people are not scared or sad.⁶¹ But people care about purpose and meaning as well, and they might sacrifice subjective well-being, for months and years, to add purpose and meaning to their lives.

With respect to subjective well-being, psychologist Daniel Kahneman and economist Alan B. Krueger have argued for a “misery index,”⁶² which might be supported on the ground that even if we cannot define happiness, and even if we are not entirely clear that happiness is something that governments should try to increase, it is surely a good idea to reduce misery. With respect to welfare, regulators might want to follow a similar strategy: we can usually specify what is bad, even if we are not entirely clear how to specify what is good. Early death, illness, sickness, accidents, and poverty are bad, and richer is better than poorer.

Of course, tradeoffs must be made, and some of them are very difficult—as, for example, when risk reduction is costly or creates substitute risks. There are also challenging questions about how to trade off emotional effects and material effects—as, for example, when an intervention makes people sad and scared, but also safer, and hence reduces mortality risks. It makes sense to say that thirty sad and anxious days are worth it if they are necessary to produce thirty additional years of life; it also makes sense to say that three thousand sad days are not worth it if they are necessary to produce only one additional day of life. The abstract idea of welfare will not inform regulators about how to make those tradeoffs. Still, it has the virtue of focusing attention on the right question.

60. For evidence, see generally Betsey Stephenson & Justin Wolfers, *Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox*, BROOKINGS PAPERS ON ECON. ACTIVITY (2008) (finding that people with better material circumstances enjoy greater subjective well-being and that increases in living standards deliver higher subjective well-being).

61. I have bracketed some complexities. A good life includes fear and sadness. Even so, regulators should try to decrease them.

62. See Daniel Kahneman & Alan B. Krueger, *Developments in the Measurement of Subjective Well-Being*, 20 J. ECON. PERSPS. 3, 4, 19 (2006).

The good news is that a great deal of research shows that it is indeed possible to measure emotional states.⁶³ Some of the relevant work relies on self-reports. We might question whether self-reported joy or anxiety actually captures internal states, but it does seem to correspond with them, at least.⁶⁴ We might also be able to measure brain activity itself.⁶⁵ Automatic facial coding can also tell us something about emotional reactions.⁶⁶ Turning emotional states into monetary equivalents remains a work in progress.⁶⁷

III. BENEVOLENT SPECTATORS

In terms of welfare and distribution, consider some possible effects of behaviorally informed interventions.

(1) An intervention would help everyone, but just by a little.

(2) An intervention would help most people by a lot, but hurt some people by a little.

(3) An intervention would help most people by a little, but hurt some people by a lot.

(4) An intervention would hurt some people by a little, but help some people by a lot.

We should be able to support the intervention in (1). Unless we introduce significant complications, we will probably want to support the intervention in (2) as well. To evaluate (3) and (4), we need to know far more, first about the relevant criteria and second about the relevant facts. With respect to criteria, it is standard to adopt a utilitarian social welfare function, purporting to treat everyone the same, in the sense

63. See generally DOLAN, *supra* note 52 (emphasizing the importance of subjective experience and focusing on pleasure and purpose).

64. See Kahneman & Krueger, *supra* note 62, at 22.

65. For one example, see generally Neil Garrett, Stephanie C. Lazzaro, Dan Ariely & Tali Sharot, *The Brain Adapts to Dishonesty*, 19 NATURE NEUROSCI. 1727 (2016) (using functional magnetic resonance imaging to show that signal reduction in the amygdala is sensitive to repeated dishonest behavior).

66. See Kate Laffan, Cass R. Sunstein & Paul Dolan, *Facing It: Assessing the Immediate Emotional Impacts of Calorie Labeling Using Automatic Facial Coding*, BEHAV. PUB. POL'Y, Nov. 11, 2021, at 13.

67. See Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 977 (2004) (noting that regulatory agencies have successfully enumerated and priced physical welfare setbacks but have largely failed to enumerate and price psychological welfare setbacks).

that everyone's welfare counts equally.⁶⁸ (I will put to one side the annoying and silly question whether it is possible to make interpersonal comparisons of utility.⁶⁹ The short and clearly correct answer is that it is.) Doing that has evident moral justifications. In one of his most remarkable passages, philosopher John Stuart Mill offered the core of the argument⁷⁰:

I must again repeat, what the assailants of utilitarianism seldom have the justice to acknowledge, that the happiness which forms the utilitarian standard of what is right in conduct, is not the agent's own happiness, but that of all concerned. As between his own happiness and that of others, utilitarianism requires him to be as strictly impartial as a disinterested and benevolent spectator. In the golden rule of Jesus of Nazareth, we read the complete spirit of the ethics of utility. To do as one would be done by, and to love one's neighbour as oneself, constitute the ideal perfection of utilitarian morality.

The central point is that utilitarianism does not play favorites. Self-dealing, on the part of the utilitarian social planner, is ruled out of bounds. In a sense, there is an overlap between the utilitarian perspective and that of philosopher John Rawls's "original position," though the latter can be seen as far more stringent.⁷¹

A. *What Is the Currency?*

Equality of treatment, understood in Mill's terms, has evident appeal. For purposes of policy, however, a crucial question is the *currency*. In this Section, I explore that question.

Are we speaking of utils? What are they, exactly? Are we speaking instead of welfare units, which might be different from utils? What conception of welfare are we using? Is it happiness or life satisfaction, or pleasure and purpose, or a bundle of incommensurable goods? Does it place a special emphasis on higher-order faculties and their exercise?

68. See generally ADLER, MEASURING SOCIAL WELFARE, *supra* note 21 (describing the utilitarian social welfare function and other frameworks for assessing government policy).

69. See generally JON ELSTER & JOHN E. ROEMER, INTERPERSONAL COMPARISONS OF WELL-BEING (1991) (arguing that it is possible to make interpersonal comparisons of well-being).

70. JOHN STUART MILL, UTILITARIANISM 24 (George Sher ed., Hackett Publ'g 2d ed. 2001) (1863).

71. See JOHN RAWLS, A THEORY OF JUSTICE 10–11 (Harv. Univ. Press rev. ed. 1999) (1971) (describing the essential features of the "original position," a hypothetical situation in which "all are similarly situated and no one is able to design principles [of justice] to favor his particular condition"). On the connection with utilitarianism, see John C. Harsanyi, *Can the Maximin Principle Serve as a Basis for Morality?*, 69 AM. POL. SCI. REV. 594, 594 (1975).

Is meaningfulness an important part of it? Are we speaking of money? Ought we to use WTP as a measure of what people care about?

It is easiest to work with WTP because often we can measure it. But money is not valuable for its own sake; it is valuable because it is a proxy for something that is. In many cases, it is a good proxy, or good enough; if people are willing to pay a great deal for something (because it brings pleasure or satisfaction, or a kind of meaning), we can have some confidence that they will benefit from it, and if they are willing to pay little or nothing, perhaps we can conclude that they will not benefit much or at all. But in some cases, money is not a good proxy. One reason is that a poor person is likely to lose more from losing a given amount of money than a rich person would, and gain more from a given unit of money than a rich person would. If you are poor, losing or gaining \$500 might mean a great deal to you; the same is not true if you are a billionaire. Poor people might be willing to spend very little for a good that would make their entire life much better; a wealthy person might be willing to spend a lot for a good that would improve just a week of their life, or perhaps just a day. The central point is that if we care about utility (or welfare), we might need to make significant adjustments in the WTP measure because it will understate the utility benefits of certain gains to poor people (or the utility costs of certain losses) and overstate the utility benefits of certain gains to wealthy people (or the utility costs of certain losses). If what matters is utility, the monetary measure will misfire.

It is nonetheless important to see that a forced exchange on terms that people dislike might make them worse off. It will require them to buy something that they do not want, undoubtedly because they want other things more. They might want to use the relevant money not to eliminate a mortality risk of 1 in 100,000 but to buy food or education or medical care, or to eliminate a mortality risk of 1 in 20,000 or 1 in 10,000. In free societies that are concerned with people's welfare, we should begin by asking people what they want, and if people do not want certain goods, we should presume that they know their own priorities. A forced exchange will decrease their welfare. Indeed, a forced exchange would violate Mill's Harm Principle without apparent justification.

A response to this objection, signaled above, is that people may suffer from a problem of "miswanting." They want some things that do not promote their welfare, and they do not want some things that would promote their welfare. In many settings, people's decisions appear not to make them happier, even when alternative decisions

would do so. Predicted welfare, or utility at the time of decision, may be very different from experienced welfare (understood as welfare as people experience it while life is actually lived). To this extent then, WTP loses much of its underlying justification. People's choices do not actually promote their welfare. If a government can be confident that people are not willing to pay for goods from which they would greatly benefit, perhaps the government should abandon WTP.

A more specific concern is that people's preferences may have adapted to existing opportunities, including deprivation. Thus Alexis de Tocqueville writes, "Should I call it a blessing of God, or a last malediction of His anger, this disposition of the soul that makes men insensible to extreme misery and often even gives them a sort of depraved taste for the cause of their afflictions?"⁷² Perhaps people show a low WTP for environmental goods, including health improvements, simply because they have adjusted to environmental harms, including those that involve health risks. Perhaps people's WTP reflects an effort to reduce cognitive dissonance through the conclusion that risks are lower than they actually are. It is not a lot of fun to think that you face serious dangers, and some people undoubtedly develop an unduly optimistic account of their actual situation. In some contexts, the idea of miswanting raises serious problems for neoclassical economics and for unambivalent enthusiasm for freedom of choice. If the basis for using WTP is welfare, there is a real difficulty: use of WTP may be imperfectly connected with promoting people's welfare.

When a behavioral market failure is involved, appropriate adjustments should be made to WTP, and the values that emerge from WTP should be corrected accordingly. It is possible, of course, that across large aggregations of workers, behavioral market failures are not a serious problem, and hence existing numbers are trustworthy. A great deal of further conceptual and empirical work needs to be done on these issues. The central point is that WTP is an effort to capture welfare effects; it is a proxy, and sometimes a poor one.

B. Prioritarianism

These are objections to the monetary measure within the welfarist framework. "Prioritarianism" suggests that we should devote special

72. ALEXIS DE TOCQUEVILLE, *DEMOCRACY IN AMERICA* 317 (George Lawrence trans., Anchor Books 1969) (1848).

attention to the welfare of those who are the least well off.⁷³ Imagine that the world consists of two people, Mary and Edna. Mary has 100 units of welfare; Edna has 1 unit of welfare. If we choose Intervention A, both will gain 20 units of welfare, so that Mary will have 120, and Edna will have 21. If we choose Intervention B, Mary will gain 10 units of welfare and Edna will gain 28, so that Mary will have 110, and Edna will have 29. Intervention A results in more aggregate welfare (141 is larger than 139), but there is an argument in favor of choosing Intervention B, not because it results in a more *equal* distribution but because (on prioritarianism grounds) it *gives more help to the person at the bottom*.

As for individuals, so for groups: we might give priority to those whose welfare is lowest and sacrifice aggregate welfare in order to achieve that goal. But how much should we sacrifice? The answer must depend on the right specification of prioritarianism. A situation in which Mary has 120 and Edna has 5 might be better than one in which Mary has 50 and Edna has 6 (or not). A situation in which Mary has 50 and Edna has 6 might be better than one in which Mary has 12 and Edna has 8 (or not).

Prioritarianism has strong defenders, though it also raises many questions, especially if it is used for regulatory purposes. One set of questions involves *incidence*: a fuel economy rule will have an assortment of effects on consumers, workers, and investors, and even if one set of effects is especially beneficial to the least well off (say, less pollution), another set of effects might be especially harmful to them (say, increased prices and fewer jobs). An occupational safety regulation might reduce risks to poor workers, but it might also result in decreases in their wages and in reduced working hours. Another set of questions cuts deeper. We might want to know why exactly Edna has 1, while Mary has 100. Did Mary work harder? Did she have better luck? Is she smarter or stronger? We might also want to know the dynamic effects of prioritarianism. If we are to prioritize the welfare of the least well off, will we reduce effort and growth?

Rawls's Difference Principle allows inequalities to the extent that they benefit the least well off.⁷⁴ The Difference Principle of course

73. See generally MATTHEW D. ADLER, *PRIORITARIANISM IN PRACTICE* (2022) (outlining the theory of prioritarianism as a branch of welfare consequentialism); Matthew D. Adler & Nils Holtug, *Prioritarianism: A Response to Critics*, 18 *POL. PHIL. & ECON.* 101 (2019) (defending prioritarianism against objections).

74. See RAWLS, *supra* note 71, at 83.

raises many questions, and it is not at all clear how to apply the Difference Principle when making regulatory choices; Rawls's own focus was on the basic structure of society, not on particular interventions. For present purposes, the only point is that prioritarianism might misfire if its use ends up having harmful aggregate effects, so that everyone ends up with 1, or perhaps 0.5.

IV. WELFARE AND DISTRIBUTION

Turn now to concrete questions, including some of the problems with which I began, and let us explore how to approach them in light of the analysis thus far. Suppose that we have an agreed-upon conception of welfare, and that we know how to operationalize it. What do we need to know to decide whether a behaviorally informed intervention increases aggregate welfare or is otherwise desirable? In some cases, it is not possible to answer that question without making contentious claims about the meaning of welfare. But in other cases, we might not be required to do that. By attending to an agreed-upon conception and issues of distributive justice, we can make a great deal of progress. We can identify paths forward.

Return to a previous example and suppose that numerous workers have been automatically enrolled in a retirement plan with a high default contribution rate (say, 7 percent). Suppose too that as a result of automatic enrollment, participation rates have become very high (say, 80 percent), and especially high among low-wage workers (say, 90 percent). As we have seen, it is possible that with automatic enrollment, people have been made worse off. Perhaps a large number of them need the money now, and perhaps they do not opt out (perhaps because of inattention and inertia). And even if most people are better off, low-wage workers might have been made worse off. They might benefit from having the money now. To be sure, the right to opt out is a significant safeguard here. But it might not be perfect. Because of inattention and inaction, they might not opt out. If not, and on certain assumptions about who is being helped and who is being hurt, we might want to adjust the automatic enrollment approach. The best solution might be a more personalized approach: automatic enrollment for most, but opt-in for some.

Or suppose that as a result of automatic enrollment in green energy, there are meaningful reductions in greenhouse gas emissions. Suppose too, however, that people end up paying significantly more for

energy, leading to welfare losses for them.⁷⁵ The environmental benefits might exceed those losses, but they might not. Without knowing the magnitude of the environmental benefits and that of the costs, we cannot know whether to celebrate or lament automatic enrollment in green energy. And even if celebration is in order, we should want to know the impact on those at the bottom of the economic ladder. If they are hurt, we might want to reassess. Again, perhaps *automatic enrollment for most, but opt-in for some*. The broader point is that whenever automatic enrollment increases participation rates across the board, there is a risk that some people will be harmed, perhaps because they do not benefit at all from participating, and perhaps because they are net losers.⁷⁶ This risk is especially troubling if an architectural nudge most harms identifiable categories of people: for example, poor people, the elderly, women, people of color, or people suffering from physical or mental disabilities.

These are points about architectural nudges and their distributional effects; something similar might be true for educative nudges. The evidence shows that in the relevant population, most people approve of mandatory calorie labels at chain restaurants.⁷⁷ But some evidence also shows that people with a good deal of self-control, and without much need to lose weight, value those labels and benefit from them—but that people lacking self-control, and with a real need to lose weight, do not value those labels at all, do not benefit from them, and in fact would prefer not to see them.⁷⁸ (There are hard normative questions here, but let us bracket them.) To that extent, it is possible that calorie labels are not justified on welfare grounds; it is also possible that such labels have undesirable distributional effects.

More generally, reminders, warnings, and disclosures of information might benefit people who do not need much help, and might not benefit, and might even hurt, the very population for whom

75. Claus Ghesla, Manuel Grieder & Renate Schubert, *Nudging the Poor and the Rich: A Field Study on the Distributional Effects of Green Electricity Defaults*, 86 ENERGY ECON. 1, 2 (2020).

76. See Jessica L. Roberts, *Nudge-Proof: Distributive Justice and the Ethics of Nudging*, 116 MICH. L. REV. 1045, 1048 (2018) (arguing that vulnerable populations may sometimes be “nudge-proof”).

77. See CASS R. SUNSTEIN & LUCIA A. REISCH, TRUSTING NUDGES: TOWARD A BILL OF RIGHTS FOR NUDGING 9–10 (2019) (providing data suggesting that a majority of Americans support a U.S. policy that mandates calorie labels at chain restaurants).

78. See Thunström, *supra* note 7, at 17–19 (describing the results of a study on the emotional taxes of calorie-salience nudges).

they are principally intended.⁷⁹ Some people do not want to know, and if they end up knowing, they lose welfare. Whenever that is so, it is quite possible that the aggregate welfare effect will be negative; it is also possible that the distributional effects are a reason for serious concern.

Or return finally to “home energy reports,” which are intended to increase energy conservation by informing people of how their energy consumption compares with that of their neighbors, and which describe various steps that customers can take to reduce their consumption.⁸⁰ Suppose that the economic savings of such nudges, when combined with the environmental benefits, greatly exceed the economic costs. If so, we might seem to have an unambiguous success story on welfare grounds. But it is easily imaginable that in the relevant population there is a great deal of heterogeneity. Some people might not like receiving home energy reports at all, notwithstanding the possibility that such reports might help them to save money; the reports might make them feel guilty, bored, or agitated. For them, home energy reports are in the nature of too much information—a point that suggests, more generally, that behaviorally informed interventions that might help people to save money, or help to reduce health or safety risks, might also have negative welfare effects. (Sometimes people would prefer not to know.)

By contrast, some people might greatly value those reports and indeed might value them in part because of their environmental benefits (and hence value them in excess of the personal economic savings they provide). Some people might be indifferent to those reports. If it turns out that many people dislike the reports and would prefer not to have them, and if those people are disproportionately poor (and would like to focus on other things or do not care about the relatively modest economic savings), the aggregate welfare effects might be negative; in any case, they are not adequately captured by the finding that the economic and environmental savings exceed the economic costs. A more targeted and personalized set of nudges—giving the reports to those who would most benefit from them—would be preferable on welfare grounds (a point to which I will return).

79. See Linda Thunström, Ben Gilbert & Chian Jones Ritten, *Nudges that Hurt Those Already Hurting – Distributional and Unintended Effects of Salience Nudges*, 153 J. ECON. BEHAV. & ORG. 267, 269 (2018) (describing how nudges may adversely impact the very groups that they were intended to aid).

80. See generally Allcott & Kessler, *supra* note 4 (conducting a study on the welfare effects of home energy reports and proposing a targeted approach to increase their benefits).

At the same time, some nudges do turn out to be more, rather than less, desirable once we disaggregate their effects. Focusing on architectural nudges, Professors Kellen Mrkva, Eric J. Johnson, and Crystal Reeck, along with PhD student Nathaniel Posner, find that “nudges, which make decisions simple or more automatic, dramatically reduce barriers that contribute to inequality. Nudges, in fact, reduced socioeconomic inequities in every type of decision we examined, including Covid-19 health decisions, retail purchases, and financial decisions.”⁸¹ Suppose that an employer or an educational institution is considering the appropriate choice architecture for decisions involving health insurance or retirement. Smart defaults, nudging people toward options that are most likely to suit their situations, might turn out to have especially large benefits for low-income choosers, for choosers with less domain-relevant expertise, or for choosers with less numeracy.⁸²

Or suppose that in order to increase participation in a program designed to provide free meals to poor children, a government agency decides automatically to enroll homeless and migrant children in that program.⁸³ On distributional grounds, such a nudge should be enthusiastically welcomed; broadly, “sludge reduction,” in the form of efforts to reduce administrative burdens and barriers, can have significant benefits for people who are old, sick, poor, or otherwise vulnerable.⁸⁴

In the same vein, some educative nudges could disproportionately benefit people at the bottom of the economic ladder, especially in light of the challenge of cognitive scarcity.⁸⁵ (Consider warnings or reminders designed to help people to avoid late fees or overuse fees, or not to miss medical appointments.) To the extent that warnings or reminders target people who are busy, occupied, or inattentive, their

81. Kellen Mrkva, Eric J. Johnson, Crystal Reeck & Nathaniel Posner, *Design Systems with Your Most Vulnerable Users in Mind*, HARV. BUS. REV. (Apr. 26, 2021), <https://hbr.org/2021/04/design-systems-with-your-most-vulnerable-users-in-mind> [<https://perma.cc/98UA-624A>].

82. Mrkva et al., *Do Nudges Reduce Disparities?*, *supra* note 23, at 67.

83. *See* Interim Rule: Direct Certification and Certification of Homeless, Migrant and Runaway Children for Free School Meals, 76 Fed. Reg. 22785 (2011).

84. *See generally* CASS R. SUNSTEIN, *SLUDGE* (2021) (discussing administrative burdens and the harms they cause).

85. *See generally* SENDHIL MULLAINATHAN & ELДАР SHAFIR, *SCARCITY: WHY HAVING TOO LITTLE MEANS SO MUCH* 39 (2013) (discussing cognitive scarcity and the challenges it poses for people who are busy, sick, elderly, hungry, lonely, or poor).

distributional effects might turn out to count in their favor.⁸⁶ Indeed, regulation of the credit card market, motivated in large part by behavioral findings, has been found to provide special help to people with poor credit ratings.⁸⁷

For some behaviorally informed interventions, a normative assessment of their distributional effects is not at all straightforward, even if the evidence is clear. For example, taxes on sugar-sweetened beverages and cigarettes might be considered regressive—a point that might be taken to count against them, even if they help to reduce both externalities and internalities (the costs that choosers impose on their future selves). But if we take account of the benefits as well as the costs of such taxes, they might not be regressive at all. It is true that if imposed on poor people, a given monetary cost produces a greater welfare loss than if imposed on wealthy people. But it is also true that the welfare benefit of such taxes might accrue mostly or disproportionately to poor people if (for example) they are the ones most likely to gain in terms of health impacts.⁸⁸ The relevant taxes might turn out to be especially beneficial, on net, to poor people.⁸⁹

Taxes are not nudges, but precisely the same conclusion might hold for architectural nudges designed to reduce consumption of high-calorie foods and cigarettes.⁹⁰ Suppose that for reasons that may or may not be clear, such interventions prove to be especially effective among low-income groups, in the sense that they produce pronounced reductions in consumption by members of such groups. It might be thought that these are regressive nudges, imposing net welfare costs on members of such groups. But a full accounting of both benefits and

86. See Fadima Bocoum, Michael Grimm, Renate Hartwig & Nathalie Zongo, *Can Information Increase the Understanding and Uptake of Insurance? Lessons from a Randomized Experiment in Rural Burkina Faso*, 220 SOC. SCI. & MED. 102, 102 (2019) (finding positive distributional effects from a behaviorally informed intervention).

87. Sumit Agarwal, Souphala Chomsisengphet, Neale Mahoney & Johannes Stroebel, *Regulating Consumer Financial Products: Evidence from Credit Cards*, 130 Q.J. ECON. 111, 162 (2015).

88. See generally Allcott et al., *Regressive Sin Taxes*, *supra* note 24 (exploring the complex distributional effects of “sin taxes,” or taxes on objects thought to be overconsumed).

89. See M. Arantxa Colchero, Juan Rivera-Dommarco, Barry M. Popkin & Shu Wen Ng, *In Mexico, Evidence of Sustained Consumer Response Two Years After Implementing a Sugar-Sweetened Beverage Tax*, 36 HEALTH AFFS. 564, 568 (2017).

90. See Lena Al-Khudairy, Samantha Johnson, Oyinlola Oyeboode, Olalekan Uthman & Rosemary Walmsley, *Choice Architecture Interventions To Improve Diet and/or Dietary Behaviour by Healthcare Staff in High-Income Countries: A Systematic Review*, 9 BMJ OPEN 1, 10, 13 (2019).

costs might suggest that these are in fact progressive nudges, with particular (net) benefits for low-income groups.

V. TARGETING AND PERSONALIZATION

The broadest lesson is that in evaluating behaviorally informed interventions, it is essential to consider their aggregate welfare effects, including their emotional impacts and their distributional effects, and to focus in particular on questions of distributive justice. In some cases, consideration of those effects and those questions will actually strengthen the argument for behaviorally informed interventions; in other cases, such consideration will severely weaken that argument; and in still other cases, such consideration will raise a set of new issues, normative and empirical.

The most interesting implication might well involve the value of shifting toward *more targeted or personalized approaches, which can often produce far higher welfare benefits than mass approaches, and which can also address distributional concerns*.⁹¹ Return to the case of home energy reports. Suppose that the population consists of three groups of people. The first group greatly values the reports; in fact, it values the reports not only because of the economic savings but also because it derives satisfaction and a kind of gratification from taking steps toward energy conservation. For members of the first group, the benefits of the reports significantly exceed the economic savings. The second group has no affective response, whether positive or negative, to the reports. The question is purely monetary: how much money can its members save? The third group does not like the report. Because of its negative reaction to the report, it would prefer not to receive it; the economic savings are outweighed by the emotional reaction. We might want to ensure that members of the first two groups receive the reports, and that members of the third group do not.

Now imagine that when people are automatically enrolled in green energy, there are two principal effects. First, the cost of energy is increased. Second, the level of pollution is decreased. We could easily

91. See, e.g., Stuart Mills, *Personalized Nudging*, 6 BEHAV. PUB. POL'Y 150, 151 (2022); Eyal Peer, Serge Egelman, Marian Harbach, Nathan Malkin, Arunesh Mathur & Alisa Frik, *Nudge Me Right: Personalizing Online Security Nudges to People's Decision-Making Styles*, 109 COMPUTS. HUM. BEHAV. 1, 2 (2020); cf. Daniel G. Goldstein, Eric J. Johnson, Andreas Herrmann & Mark Heitmann, *Nudge Your Customers Toward Better Choices*, HARV. BUS. REV. (Dec. 2008) (calling for private companies to begin nudging consumers toward options that benefit "both the company's and customers' long-term interests").

imagine communities, or even nations, for which automatic enrollment is justified on welfare grounds; we could also imagine communities or nations in which it is not so justified. Suppose that it is justified on those grounds. It might nonetheless be the case that the higher cost hurts people at the bottom of the economic ladder. If so, we might favor opt-out for the majority and opt-in for those at the bottom, at least if we do not have reason to believe that the adverse environmental effects make that outcome intolerable. Of course, it would also be possible to adopt a system of automatic enrollment for all, alongside subsidies for those at the bottom of the economic ladder.

Could educative nudges be treated similarly? At first glance, they certainly could be. For example, *calorie labels for thee, but not for me*. Imagine a situation in which calorie labels were provided only to those individuals, and those groups, that would benefit from them. People might benefit because they would make healthier choices as a result, because they would enjoy learning about the caloric content of food, or both. People might not benefit, and might be worse off, if they would not make healthier choices as a result and if they would simply enjoy their meals less. If private or public institutions could learn who falls in what category, they could personalize their calorie labels accordingly.

Of course, it might not be feasible to do that. But in the fullness of time, technology will make personalization increasingly simple. The result could be maximization of social welfare and promotion of distributive justice.

But let us return to the largest issues. We cannot achieve those results without asking the questions with which I began: (1) *What are the aggregate effects on social welfare?*; (2) *What are the effects of the intervention on people's emotional states?*; (3) *Who is likely to be helped and who is likely to be hurt?*; (4) *What are the expected effects on the least well off?*; and (5) *Do the benefits to those who are helped exceed the costs to those who are hurt?* As (2) emphasizes, we should focus not only on the material effects of interventions but also on how interventions make people *feel*. If they make people feel scared or sad, or resentful or humiliated, they might nonetheless be worthwhile on balance, but perhaps they can be designed in a way that will help people feel happier or in control. Regulators should attempt to do that. If they succeed, interventions might well be more likely to work, and less likely to trigger reactance. In any case, they will be kinder.

If we aim to make the world a more “pleasant and exciting place,”⁹² the five questions should be the foundation not only for evaluation of behaviorally informed regulation but also for evaluation of regulation in general.

CONCLUSION

It is standard to evaluate behaviorally informed interventions, including nudges, by asking if they are cost-effective.⁹³ If they can achieve a given end at a cost of \$2 million rather than \$10 million, who could possibly object? It is tempting to evaluate such interventions by looking at effects on participation rates. If an intervention increases participation from 20 percent to 70 percent, do we not have a significant success story? The problem is that cost-effectiveness is not the same as a welfare improvement; we need to know whether achievement of the relevant end justifies the expenditure. We also need to know what, exactly, people are gaining from participating in some program or opportunity. Perhaps they are gaining very little; perhaps they are losing.

My minimal submission here, and I hope the least contentious, has been that the welfare question is the right one. I have urged, more specifically, that it is essential to focus on questions of distributive justice. If an intervention does not help the most vulnerable, it might be redirected and improved. I have also emphasized the importance of considering the effects of interventions on people’s emotional states; the argument for attention to those effects might well be counted as the beating heart of the argument here. Do such interventions make people feel sad or frightened? Do they alleviate or increase anxiety? Do they give people a sense of confidence? Do they produce a nod or a smile, or instead a grimace? These questions matter in part because an intervention will often be more effective if it does not make people feel terrible or miserable. But they also matter because emotional states are important in themselves. Attending to those states is a crucial part of the welfare analysis. It is also the right thing to do, for one reason above all: it is kind.

92. RAMSEY, *supra* note 1, at 291.

93. Benartzi et al., *supra* note 49, at 1042.