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

HARVARD

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

BACK TO MILL? BEHAVIORAL WELFARE ECONOMICS

Cass R. Sunstein

Discussion Paper No. 1026

01/2020

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center

The Social Science Research Network Electronic Paper Collection:
https://ssrn.com/abstract=3499180

Very preliminary draft 1/7/20
All rights reserved

Back to Mill? Behavioral Welfare Economics

Cass R. Sunstein*

Abstract

A growing body of normative work, going under the name of "behavioral welfare economics," explores how deference to people's choices might be reconciled with behavioral findings about human error. This work has important implication for economic analysis of law. The best approach adopts a working presumption in favor of respect for those choices, so long as they are adequately informed and sufficiently free from behavioral biases. For purposes of law, it is most helpful to emphasize that people may choose the wrong means to promote their own ends; in principle, that possibility may legitimate nudges, incentives, or mandates. But for both empirical and philosophical reasons, those interested in law and policy, or in welfare itself, should be cautious about making strong normative claims about whether people choose the right ends. Behavioral welfare economics, and behaviorally informed analysts of law, should be committed to the working presumption, with humility and aware of the inevitability of taking a stand on some fundamental philosophical issues.

I. The Goal

Much of behavioral economics, and much of behaviorally informed analysis of law, focuses on departures from standard accounts of rationality.[1] Exploring actual behavior, it seeks to avoid the most contentious normative questions about the relationship between choice and welfare. But an illuminating and growing body of work explores whether and in what sense economists, lawyers, and others interested in behaviorally informed public policy can continue to insist on the sovereignty of individual preferences, while also acknowledging behavioral findings.[2] My central aim here is to argue for a degree of humility, captured in a working presumption in favor

* Robert Walmsley University Professor, Harvard University. I am grateful to Hunt Allcott and David Strauss for valuable comments on a previous draft.

[1] For a clear catalogue, see Richard Thaler, Misbehaving (2015).

[2] See B. Douglas Bernheim and Antonio Rangel, Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics, 124 Q J Econ 51(2009); B. Douglas Bernheim and Antonio Rangel, Toward Choice-Theoretic Foundations for Behavioral Welfare Economics, 97 Am Econ Rev: Papers and Proceedings 464 (2007); R. Douglas Bernheim, Behavioral Welfare Economics, 7 J European Economic Assn. 267 (2009); Hunt Allcott and Judd Kessler, The Welfare Effects of Nudges: 11 Am Econ Journal: Applied Economics 263 (2019); Linda Thunstrom, *Welfare Effects of Nudges: The Emotional Tax of Calorie Menu Labeling*, 14 Judgment and Decision Making 11 (2019); Hunt Allcott and Cass R. Sunstein, Regulating Internalities, 34 Journal of Policy Analysis and Management 698 (2015).

of respect for people's informed and behaviorally unbiased judgments.[3] As we shall see, this presumption should be accompanied by a distinction between "means paternalism" and "ends paternalism," of special relevance to behaviorally informed law and policy.

More specifically, I shall offer two arguments. The first is that it is productive to ask whether people are choosing the wrong means to promote their own ends. Asking that question can illuminate constructive inquiries into appropriate regulatory responses, which might include mandates, incentives, or nudges.[4] The second is that behavioral economists, and those interested in behaviorally informed approaches to law, should be cautious about making strong normative claims about whether people choose the right ends. They approach that question with humility, and not only for empirical reasons: The ends that people choose might make their lives go less well. Nonetheless, the working presumption, emphasizing adequate information and an absence of behavioral biases, provides a reasonable path forward, with many implications for research, policy, and law.[5]

II. Choices and Welfare

Do people's choices promote their welfare? For reasons that are now well-known, the best answer is: Sometimes. If people lack information, they might choose poorly.[6] If sellers exploit people's behavioral biases, consumers' choices may go awry.[7] Even without self-conscious exploitation of such biases, people may choose poorly because of limited attention, inertia, present bias,[8] hyperbolic discounting, loss aversion, or unrealistic optimism,[9] potentially justifying a regulatory response.[10] People may make mistaken predictions about the effects of options on their welfare.[11] It is also true that people are susceptible to supposedly irrelevant factors or "frames"[12]: In an opt-in system, for example, they might end up in a very different

[3] Allcott and Sunstein, supra note, is in this vein. The presumption could be fortified with reference to the potentially self-interested or malevolent incentives of those who seek to interfere with people's judgments. I am bracketing that important point here.
[4] See Jacob Goldin and Nicholas Lawson, Defaults, Mandates, and Taxes, 18 American Econ Review 438 (2016).
[5] Endogenous or shifting tastes raise distinctive challenges, taken up in various places below. See, e.g., Linus Mattauch and Cameron Hepburn, Climate Policy When Preferences Are Endogenous—and Sometimes They Are, 40 Midwest Studies in Philosophy 76 (2016); Jon Elster, Sour Grapes (1983); Menahem Yaari, Endogenous Changes in Tastes: A Philosophical Discussion, in Decision Theory and Social Ethics 56 (Hans Gottinger and Werner Leinfellner eds. 1976).; Brent McConnell, Endogenous Preferences and Welfare Analysis (2006), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=933089
[6] See Oren Bar-Gill, Seduction by Contract (2012).
[7] See id.; George Akerlof and Robert Shiller, Phishing for Phools (2016).
[8] See Ted O'Donogue and Matthew Rabin, Present Bias, 105 Am Econ Rev 273 (2015).
[9] See Xavier Gabaix, Behavioral Inattention (2019), available at https://www.nber.org/papers/w24096; Natasha Sarin, Making Consumer Finance Work, 119 Colum L Rev 1519 (2019).
[10] See Bar-Gill, supra note; Sarin, supra note; Sarah Conly, Against Autonomy (2002).
[11] See Cass R. Sunstein, Ruining Popcorn? The Welfare Effects of Information, 58 J Risk & Uncertainty 121 (2019).
[12] See Perspectives on Framing (Gideon Keren ed. 2011); Marc Scholten et al., The Framing of Nothing and the Psychology of Choice, J Risk and Uncertainty (2019), available at https://link.springer.com/content/pdf/10.1007%2Fs11166-019-09313-5.pdf

situation from where they end up in an opt-out system.[13] If so, what is their preference[14]? People's preferences may be endogenous to legal rules[15]; whether or not that it is, preferences may shift over time, and choosers may not appreciate that fact.[16] These questions raise challenges for efforts to base law and policy on people's preferences, or to conduct any kind of welfare analysis on the basis of preferences.[17] They help explain the growing effort in multiple fields to investigate subjective well-being, entirely unanchored in preferences.[18]

Some of the most careful and illuminating discussions of the underlying issues, not yet engaged in law and policy, come from by R. Douglas Bernheim,[19] who begins by noting that "standard welfare economics" associates welfare with choices. We can understand Bernheim's work as the best and clearest effort to rescue the foundations of standard economic theory while acknowledging behavioral findings. It also bears directly on legal questions, such as the domain of paternalism, and for that reason, it is worth careful attention here. Bernstein suggests that standard welfare economics invokes three general premises[20]:

> *Premise 1: Each of us is the best judge of our own well-being.*
> *Premise 2: Our judgments are governed by coherent, stable preferences.*
> *Premise 3: Our preferences guide our choices: when we choose, we seek to benefit ourselves.*

Something like these premises played a central role in the early decades of economic analysis of law,[21] and it continues to play a significant role today.[22] Bernheim recognizes that the

---

[13] See Jon Jachimowitz et al., Why and When Defaults Influence Decisions: A Meta-Analysis of Default Effects, 3 Behavioural Public Policy 159 (2019).

[14] For valuable discussion, see Jacob Goldin, Which Way to Nudge? Uncovering Preferences in the Behavioral Age, 125 Yale L.J. 1 (2015); Jacob Goldin, Libertarian Quasi-Paternalism, 82 Missouri L. Rev. 669 (2017). To summarize a complex argument, Goldin argues for a distinction between inconsistent choosers, whose choice depends on the default, and consistent choosers, whose choice is impervious to the default. In his account, it is "quasi-paternalistic" to nudge inconsistent choosers. That is a helpful distinction. At the same time, consistent choosers might be impairing their own welfare, perhaps because of present bias or unrealistic optimism.

[15] Linus Mattauch and Cameron Hepburn, Climate Policy When Preferences Are Endogenous—and Sometimes They Are, 40 Midwest Studies in Philosophy 76 (2016).

[16] See Edna Ullmann-Margalit, Big Decisions: Opting, Converting, and Drifting, 58 Royal Institute of Philosophy Supplement 157 (2006).

[17] For different efforts to meet those challenges, see Richard H. Thaler and Cass R. Sunstein, Nudge (2009); Jacob Goldin, Which Way to Nudge? Uncovering Preferences in the Behavioral Age, 125 Yale L.J. 1 (2015).

[18] See, e.g., Paul Dolan, Happiness By Design (2014); Daniel Kahneman et al., Back to Bentham? Explorations of Experienced Utility, 112 Q J Econ 375 (1997); David Schkade and Daniel Kahneman, *Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction*, 9 Psych. Science 340 (1996). For valuable discussion of foundational issues, see Matthew Adler, Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis (2011). On law and policy, see John Bronsteen et al., Well-Being Analysis vs. Cost-Benefit Analysis, 62 *Duke Law Journal* 1603 (2013).

[19] See R. Douglas Bernheim, The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics, 7 J Benefit-Cost Analysis 12 (2016).

[20] Bernheim, supra note, at 16.

[21] It is implicit, for example, in Richard Posner, Economic Analysis of Law (1973).

[22] For one example in a specific context, see Ted Gayer and W. Kip Viscusi, Overriding Consumer Preferences with Energy Regulations, 43 J Reg Econ 248 (2013). For a general account, see Joshua D. Wright and Douglas H.

resulting understanding "may fall short of a philosophical ideal," but urges that this "should not trouble us excessively," because it "captures important aspects of well-being and lends itself to useful implementation."[23] That is hardly an unreasonable claim, but it leaves open questions and it remains to be specified; it might also point in the wrong directions. Consider, for example, the question whether a soda tax is a good idea[24]; whether energy efficiency regulations can be justified as a response to consumer mistakes[25]; whether heavy taxes on cigarettes might make smokers better off[26]; whether motor vehicle safety regulation might provide people with "experience goods," such as rear visibility, and thus improve their lives by their own lights.[27] In these and other cases, falling short of a philosophical ideal might turn out to be a fatal flaw.

And what, exactly, is meant by the claim that the standard understanding "captures important aspects of well-being?" Perhaps the suggestion, empirical in nature, is that much of the time, satisfaction of people's preferences does, as a matter of fact, promote their well-being, simply because they know what they like.[28] Bernheim makes something like this claim, invoking "the central Cartesian principle that subjective experience is inherently private and not directly observable."[29] He adds: "We know how we feel; others can only make educated guesses. These considerations create a strong presumption in favor of deference to our judgments."[30]

But Bernheim also makes a separate argument, involving autonomy rather than welfare: "my views about my life are paramount because it is, after all, *my* life."[31] This is a Kantian idea, suggesting the importance of respect for choosers, even if they err. People are ends, not means. Writing in this vein, Jeremy Waldron urges: "Deeper even than this is a prickly concern about dignity. What becomes of the self-respect we invest in our own willed actions, flawed and misguided though they often are, when so many of our choices are manipulated to promote what someone else sees (perhaps rightly) as our best interest? . . . I mean dignity in the sense of self-respect, an individual's awareness of her own worth as a chooser."[32] An emphasis on the idea that people are making choices about *their* lives might reflect a commitment to respect for

---

Ginsburg, *Behavioral Law and Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 Nw. U. L. Rev. 1033 (2015).

[23] Bernheim, supra note, at 16.

[24] See Hunt Allcott et al., Should We Tax Sugar-Sweetened Beverages?, 33 J Econ Persp 202 (2019).

[25] Hunt Allcott, Paternalism and Energy Efficiency: An Overview, 8 Annual Review of Economics 145 (2016).

[26] Jonathan Gruber and Sendhil Mullainathan, Do Cigarettes Taxes Make Smokers Better Off, 5 B.E. Journal of Economic Analysis and Policy (2005), available at https://www.degruyter.com/view/j/bejeap.2005.5.issue-1/bejeap.2005.5.1.1412/bejeap.2005.5.1.1412.xml?lang=en

[27] See Cass R. Sunstein, Rear Visibility and Some Unresolved Problems for Economic Analysis, 10 J of Benefit-Cost Analysis 317 (2019).

[28] For a series of objections, see Sarah Conly, Against Autonomy (2012).

[29] Id. at 17.

[30] Id.

[31] Id.

[32] See Jeremy Waldron, It's All For Your Own Good, NY Rev of Books (2014), available at http://www.nybooks.com/articles/archives/2014/oct/09/cass-sunstein-its-all-your-own-good/

dignity and autonomy, not welfare at all.[33] That form of respect might stand as a decisive objection to paternalism in law.[34]

These arguments overlap with those in the canonical text on these questions, *On Liberty*, where Mill insisted[35]:

> The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right.

Mill offered a number of separate justifications for his famous Harm Principle, but his most emphatic, and the most relevant here, is epistemic. It is in the same family as the first of Bernheim's: choosers are in the best position to know what is good for them. In Mill's view, the problem with outsiders, including government officials, is that they lack the necessary information. 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."[36] 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."[37] If the goal is to ensure that people's lives go well, Mill concludes that the best solution is to allow people to find their own path. Compare F.A. Hayek's remarkable suggestion that "the awareness of our irremediable ignorance of most of what is known to somebody [who is a chooser] is *the chief basis of the argument for liberty*."[38] For Hayek, the key contrast is between the chooser, who knows a great deal, and the outsiders, who show "irremediable ignorance," especially if they are social planners.

Important strands in contemporary behavioral welfare economics make elaborate and instructive arguments in Hayek's direction, suggesting that deference to individual choice is the right approach, subject to specific qualifications.[39] Bernheim's emphasis on "the central

---

[33] Some questions about this idea are raised in Cass R. Sunstein, Is Deontology A Heuristic?, 63 Iyyun: The Jerusalem Philosophical Quarterly 83 (2014).

[34] For a defense of coercive paternalism, see Ryan Bubb and Richard Pildes, How Behavioral Economics Trims Its Sails and Why, 127 Harv L Rev 1593 (2014). By design, Bubb and Pildes do not engage this kind of normative objection: "We thus put aside various external critiques that could be, and have been, mounted. Autonomy- or liberty-based political theories argue, for example, that these values should have priority over welfare maximization." Id. at 1601. The only point is that if a Kantian idea, of the sort invoked by Bernheim and Waldron, is in play, it would serve as a check on imaginable claims in behaviorally informed law and policy.

[35] John Stuart Mill, On Liberty (2d ed. 1863), reprinted in The Basic Writings of John Stuart Mill: On Liberty, The Subjection of Women, and Utilitarianism 3, 11-12 (Dale E. Miller ed., 2002).

[36] Id.

[37] Id.

[38] Friedrich Hayek, The Market and Other Orders, in The Collected Works of F. A. Hayek 384 (Bruce Caldwell ed., 2013).

[39] E. Douglas Bernheim and Dmitry Taubinsky, Behavioral Public Economics, in Handbook of Behavioral Economics: Foundation and Applications 1 381 (R. Douglas Bernheim et al. eds. 2018), can easily be read in this way. For an influential early effort, see Edward Glaeser, Paternalism and Policy, 73 U. Chi. L. Rev. 133 (2006).

Cartesian principle that subjective experience is inherently private and not directly observable," and on our unique knowledge of "how we feel," fits well with Mill's claim that the "ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else." When people choose chocolate over vanilla, salmon over tuna, basketball over football, rest over recreation, or spending over saving, they do so because they know what they like. Outsiders are most unlikely to have that knowledge.[40]

To be sure, we should be cautious here. It is true that even if people usually have unique knowledge of "how they feel," their knowledge on that count may not be perfect.[41] In some cases, external observers might know better, especially if the area requires technical expertise.[42] Recall that choosers must solve a prediction problem; they must decide, at some point in advance of actual experience, about the effects of one or another option on experience.[43] To solve that problem, knowing "how they feel" is not enough. At a minimum, they must know "how they *will* feel," and they might not know nearly enough to know that. But as a general rule, the claim for the epistemic advantages of choosers is more than plausible. At the same time, it has taken a real battering from behavioral findings.[44]

III. Direct and Indirect Judgments

Many people are troubled by that apparent battering.[45] For current purposes, the most important challenge involves Premise 1. Behavioral findings seem to suggest that "people do not reliably exercise good judgment." More precisely, behavioral science shows that people sometimes make mistakes about what would promote their own well-being, and in some cases, they are not the best judges of what would do that.[46] If so, Mill's epistemic argument is severely

---

[40] A rejoinder can be found in Conly, supra note, at 41: "We don't regard it as insulting to assume that the man on the street can't do quantum mechanics, because he can't (unless you're on a very special street). The paternalist believes that it is the facts that suggest a chance in the status we accord people, achage from what we might have thought about ourselves to a more realistic acceptance of our inabilities. The suggestion here is simply that we should treat people in accordance with their real abilities and their real limitations."

[41] See Timothy Wilson, Strangers to Ourselves (2004).

[42] See Ryan Bubb and Richard Pildes, How Behavioral Economics Trims its Sails and Why, 127 Harv L Rev 1593 (2014).

[43] See Daniel Kahneman et al., Back to Bentham? Explorations of Experienced Utility, 112 Q. J. Econ. 375 (1997). Sunstein, Rear Visibility, supra note, finds what seems to be a significant mistake, on the part of consumers, with respect to a technology designed to make driving safer.

[44] An ambitious account is Sarah Conly, Against Autonomy (2012); in the same vein, see Ryan Bubb and Richard Pildes, How Behavioral Economics Trims its Sails and Why, 127 Harv L Rev 1593 (2014).

[45] For different perspectives, see Goldin, Libertarian Quasi-Paternalism, supra note; Bernheim and Taubinsky, supra note; Allcott and Sunstein, supra note.

[46] For some vivid examples, see Saurabh Bhargava et al., Choose to Lose, 132 Q J Econ 1319 (2017), available at https://academic.oup.com/qje/article-abstract/132/3/1319/3769420; Saurabh Bhargava et al., Do Individuals Make Sensible Health Insurance Decisions? Evidence from a Menu with Dominated Options (2015), available at https://www.nber.org/papers/w21160; Jason Abaluck and Jonathan Gruber, Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance (2013) https://www.nber.org/papers/w19163; Christopher Afendulis et al., Dominated Choices and Medicare Advantage Enrollment, 20 J Econ Behav Organ. 72 (2015); Jason Abaluck and Jonathan Gruber, Choice Inconsistencies Among the Elderly (2009), available at https://www.nber.org/papers/w14759;

undermined. For purposes of law and policy, the door would seem open to paternalism, not only in the form of nudges, but also in the form of mandates and bans.[47]

In what might be taken as an effort to rescue the essentials of standard welfare economics and Mill's basic account, Bernheim responds is that "the argument is faulty because it conflates what I will call direct and indirect judgments."[48] A direct judgment involves ultimate objectives, or outcomes that people care about for their own sake. An indirect judgment involves alternatives that lead to those outcomes. We might understand direct judgments as involving intrinsic goods, and indirect judgments as involving goods that are instrumental to their realization. As an example, Bernheim points to a person, called Norma, who is choosing between one boxes, a red one containing a pear and a yellow one containing an apple. She prefers apples, but she chooses the red one, mistakenly ending up with a pear. This is a welfare-reducing indirect judgment, and Bernheim agrees that Norma has made a mistake: "Behavioral economics and psychology provide us with ample reason to question certain types of indirect judgments."[49] In the face of a mistaken indirect judgment about money, safety, or health, a response might take the form of mandatory disclosure, or a warning, or perhaps some kind of default rule; in extreme cases, it might justify a ban.[50] But so long as we are speaking only of indirect judgments, we might insist that we have not departed radically from Mill's general framework, amending it only to say that for indirect judgments, his epistemic argument sometimes fails.

We can link Bernheim's argument here with the suggestion that behavioral economics justify "means paternalism" but not necessarily "ends paternalism."[51] The basic idea is that people are sometimes mistaken about how to get to their own preferred destination. On that view, behaviorally informed interventions increase *navigability*. A GPS device is a form of means paternalism; it allows drivers to specify where they want to go (and helps them to get there). A default rule, automatically enrolling people in some program, can be seen as means paternalistic insofar as it is thought that (most) people prefer to be enrolled in that program (even if they suffer from inertia).[52] A calorie label can be seen as a form of means paternalism insofar as it enables people to make their own choices about how to promote their ends.

Many efforts to increase navigability, embodying a form of means paternalism, retain freedom of choice, and so can be seen as compatible with Mill's Harm Principle.[53] But some do not. A ban on trans fats can be regarded as a form of means paternalism, at least if we are clear that the ban fits with, and does not undermine, people's ends.[54] Occupational safety requirements can be seen in similar terms, even if it overrides the choices of unrealistically optimistic workers who would be willing to run the risks that those requirements eliminate.[55] When behavioral

[47] See Conly, supra note; Bubb and Pildes, supra note.
[48] Bernheim, supra note, at 17.
[49] Id. at 18.
[50] See Conly, supra note.
[51] See id.; Cass R. Sunstein, Why Nudge? (2014).
[52] See Brigitte Madrian and Dennis O'Shea, The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior, 117 Q J Econ 377 (2002).
[53] See Cass R. Sunstein, On Freedom (2019).
[54] See Conly, supra note.
[55] See George Akerlof and William Dickens, The Economic Consequences of Cognitive Dissonance, 72 Am Econ Rev 307 (1982).

economics suggests that people are making a mistake, we have a *behavioral market failure*, in the sense that people's judgments lead to some kind of welfare loss, because of an absence of available information or the presence of a behavioral bias.[56] Much of behavioral law and economics is focused on that problem.[57] As invoked for purposes of law and policy, behavioral market failures typically involve means paternalism.[58] To see the underlying issues, let us use Bernheim's terminology, distinguishing between direct and indirect judgments.

Does that distinction work? An initial problem is that if Norma prefers apples to pears, it is natural to ask whether an indirect judgment is involved as well. Apples are good, but it would be hard to justify the conclusion that they are "ultimate objectives, or outcomes that people care about for their own sake." (Friendship might be an intrinsic good; not so much apples.) If Norma prefers apples to pears, it is probably because she thinks that they taste better (or perhaps that they are healthier). The legal applications are numerous, including regulatory responses to unhealthy eating,[59] insufficient savings,[60] and "dark patterns" online.[61] Bernheim is alert to this point and adds[62]:

> Now let us add a wrinkle: assume Norma's ultimate goal is to achieve certain mental states ("internal goods"). From that perspective, all consumption items ("external goods") are means to ends, and choices among them always involve indirect judgments. Moreover, just as Norma may misjudge the contents of a box, she may also misapprehend the relationships between consumption goods and mental states. *However, assuming she is sufficiently familiar with apples, pears, and bananas to understand the consequences of eating each, her indirect judgments among open boxes will be correctly informed, and hence will faithfully reflect her direct judgments.*

What is rightly added here is more than a "wrinkle"; it is fundamental to behaviorally informed law and policy, and also to behavioral welfare economics. Norma is probably not concerned *only* with internal mental states (she might well care about price, health, and morality), but when it comes to food choices, her mental state is almost certainly something that she cares about. Consumption choices are typically means to ends, and in that sense they typically involve indirect judgments. Moreover, the assumption of sufficient familiarity might not turn out to hold. For many choices, people are not sufficiently familiar with the options "to understand the consequences" of each, and even if they are, they might suffer from some kind of behavioral bias, such as present bias.[63] In this light, we should add that for Norma's choice of apples to be a good measure of her welfare, she must not only be informed ("sufficiently familiar") but also free from any such bias. Modifying Mill, we might adopt a working

---

[56] See George Akerlof and Robert Shiller, Phishing for Phools (2016).
[57] See Bar-Gill, supra note; Bubb and Pildes, supra note.
[58] See Bar-Gill, supra note.
[59] See Matthew Rabin, Healthy Habits: Some Thoughts on The Role Of Public Policy in Healthful Eating and Exercise Under Limited Rationality, in Behavioural Public Policy (Adam Oliver ed. 2013).
[60] See Bubb and Pildes, supra note.
[61] See James Ligori and Lior Strahilevitz, Shining A Light on Dark Patterns (2019), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3431205
[62] Bernheim, supra note, at 18.
[63] Cigarette smoking is an example. See Marianna Masiero, Personal Fable: Optimistic Bias in Cigarette Smokers, Int. J. High Risk Behav. Addict 20939e (2015).

presumption in favor of this amended version of Premise 1, designed to orient behavioral welfare economics as applied to law and policy:

> *Working presumption: Each of us should be taken by outsiders to be the best judge of our own well-being, to the extent that we are adequately informed and sufficiently free of behavioral biases.*

Note that the working presumption does not depend on a distinction between direct and indirect judgments, or ends and means, though it is true that in application, it is usually likely to support interventions in the interest of means paternalism. Note also that the words "should be taken by outsiders to be," as distinguished from the more dogmatic "are"; I return to the difference below. And of course there are questions about how to operationalize the working presumption. It would make sense to examine what choices people make when they are actually well-informed; to see what choices people make when they do not suffer from limited attention and are in a position to evaluate all relevant facets of an option; to use people's active choices rather than passive ones, which may be a product of inertia; and to use otherwise unbiased choices, such as long-run choices based on a realistic understanding of facts, rather than biased ones, such as those that reflect present bias or optimistic bias.[64] Ideas of this kind can be seen as an effort to draw on a broadly Millian understanding, respectful of private choices, while also recognizing information deficits and behavioral biases.[65] They might provide a way to discipline behavioral welfare economics in areas that include savings behavior, decisions with respect to energy-efficient products, and choices of high-calorie or low-calorie food.[66]

By contrast, one of Bernheim's goals is to insist on respect for direct judgments. Instead of the working presumption immediately above, he argues in favor of two premises[67]:

> *Premise A: With respect to matters involving either direct judgment or correctly informed indirect judgment, each of us is the best arbiter of our own well-being.*

> *Premise B: When we choose, we seek to benefit ourselves by selecting the alternative that, in our judgment, is most conducive to our well-being.*

In defense of Premise A, he urges that existing economic research does not "provide evidence that people exercise poor direct judgment – for example, that they like certain goods or experiences 'too much' and others 'not enough.'"[68] On the contrary, he maintains, "[t]he occasional objection to a direct judgment entails nothing more than a difference of opinion between the analyst and the consumer as to what constitutes a good or fulfilling life."[69] Those are strong words, and they collapse the distinction between two very different questions: (1) Do people know what they will like, before they have it? (2) Do people know what they like, when

---

[64] See Allcott and Sunstein, supra note.
[65] Akerlof and Shiller, supra note, is in the same vein, though it emphasizes the active efforts by sellers to exploit those deficits and biases, or to "phish for" them.
[66] See Linda Thunstrom, Welfare Effects of Nudges, 14 Judgment and Decision Making 11 (2019).
[67] Bernheim and Taubinsky, supra note, at 16.
[68] Bernheim, supra note.
[69] Id.

they are having it? Repeating that sentence elsewhere, Bernheim and Taubinsky add: "Thus there is no objective foundation for overturning the presumption in favor of a direct judgment and declaring the analyst's perspective superior."[70]

If all we have is a "difference of opinion" between the analyst and the chooser, we do not, by stipulation, have an "objective foundation" for favoring the views of the analyst. But what if the analyst has data[71]? With respect to question (1) in the immediately preceding paragraph, what if the analyst has information about what people are likely to like, and what if that information suggests that people's ex ante predictions are incorrect? With respect to question (2), what if the analyst has an account of what makes for a good or fulfilling life, and not simply an "opinion"[72]?

IV. Practice and Theory

I now offer two claims. The first is that for behavioral economics and behaviorally informed law and policy, indirect judgments, or judgments about means, really are the coin of the realm. If we are speaking about inertia, present bias, unrealistic optimism, probability neglect,[73] or limited attention, we are almost always dealing with judgments or decisions that might defeat people's own ends. And if we are speaking of default rules, disclosure, reminders, warnings, or uses of social norms, we are almost always dealing with efforts to encourage people to choose better means to achieve their own ends. Behavioral welfare economics typically deals with indirect judgments, and if it embraces paternalism, it is means paternalism.

The second is that with respect to direct judgments, behavioral welfare economics, like standard welfare economics, should proceed with humility. One reason is empirical; another is normative. To say that informed choosers are the best arbiters of their own well-being is to take a contested stand on how to think about the very idea of well-being. In the liberal tradition, time-honored ideas about autonomy and welfare do support a working presumption in favor of something like that proposition. But the presumption should be embraced with caution, and with an understanding of what kind of stand it is taking.

A. The Pervasiveness of Indirect Judgments

As applied to law and policy,[74] almost all of behavioral economics involves indirect judgments. The bulk of behavioral research involves such judgments as well.[75] For example, the choice of one product over another – say, Toyota RAV4 over the Toyota RAV4 Hybrid – will

---

[70] See Bernheim and Taubinsky, supra note, at 16.

[71] See Steven Levitt, Heads or Tails: The Impact of A Coin Toss on Major Life Decisions and Subsequent Happiness (2016), available at https://www.nber.org/papers/w22487.

[72] See Fred Feldman, What Is This Thing Called Happiness? (2010).

[73] See Cass R. Sunstein, Probability Neglect, 112 Yale L.J. 61 (2002).

[74] See Sendhil Mullainathan and Eldar Shafir, Scarcity (2013). We can understand policy to include some behaviorally informed judgments of private institutions. See David Laibson, Private Paternalism, the Commitment Problem, and Model-Free Equilibrium, 108 Am Econ Rev: Papers and Proceedings 1 (2018).

[75] See, e.g., Jeffrey Kling et al., Comparison Friction: Experimental Evidence from Medicare Drug Plans, 127 Q J Econ 199 (2012).

almost certainly be a result of a number of subsidiary judgments: how the preferred vehicle looks, how it drives, how reliable it will be, how big it is, the storage space, the purchase price, the anticipated cost of operation, how often one will have to go to a gas station and refuel. What is the direct judgment here? We should agree that if a consumer named Jane makes some kind of clear mistake – she wants a smaller car and wrongly believes that the hybrid model is bigger – she is like Norma, selecting the wrong box. But suppose that Jane make a subtler error: She is insufficiently responsive to fuel savings, in the sense that she would have saved a lot of money with the RAV4 Hybrid, but that because of present bias, she decided against it.[76] Unless we introduce other considerations (such as a liquidity constraint), Jane was mistaken. She was not, in this case, the best judge of her own well-being. Some kind of intervention would appear to be necessary, perhaps in the form of clear disclosure of relevant information.

Perhaps we should say that Jane was not "sufficiently familiar with" the two kinds of cars "to understand the consequences" of choosing one or the other, and so her indirect judgments between the two was not "correctly informed, and hence" did not "faithfully reflect her direct judgments." Fair enough. But if that is the case, how much remains of the idea of deference to individual judgment in cases of genuine or realistic interest to law and public policy[77]? A sensible working presumption, described above, is that in many circumstances, people are good indirect choosers, likely better than anyone else, but that when they lack information or suffer from a behavioral bias, their indirect choices will not be reliable. Whether particular choices fall in that category is an empirical question.

To broaden the viewscreen: Many laws and regulations involves externalities, and for that reason, much of the welfare analysis will not be distinctly behavioral. For example, fuel economy regulations impose costs, which can be calculated (and which are imposed largely on consumers, in the form of higher sticker prices).[78] Those regulations also reduce air pollution, including greenhouse gas emissions; monetization is more challenging here, but standard tools are available to do exactly that.[79] To this extent, the welfare analysis is conventional.[80] The behavioral dimension comes from savings to consumers, particularly through economic savings (from reduced gas usage) and also through time savings (from fewer visits to the gas station).[81]

[76] See Kenneth Gillingham et al., Consumer Myopia in Vehicle Purchases (2019), available at https://www.nber.org/papers/w25845.

[77] Recall Bernheim's acknowledgment of the possibility that "all consumption items ('external goods') are means to ends, and choices among them always involve indirect judgments." It is true that some governments impose coercive controls on intimate aspects of people's lives -- as, for example, by criminalizing same-sex relations – and these controls might, at some point, be defended on behavioral grounds. But in the institutions that are now using behavioral economics, indirect judgments are the lay of the land, and the same is true of academic research that finds mistakes or recommends a behaviorally informed intervention.

[78] See Antonio Bento et al., Estimating the Costs and Benefits of Fuel Economy Standards (2019), available at https://www.nber.org/papers/w26309

[79] See Michael Greenstone, Developing a Social Cost of Carbon for US Regulatory Analysis, 7 Review of Environmental Economics and Policy 23 (2013).

[80] Ted Gayer and W. Kip Viscusi, Overriding Consumer Preferences With Energy Regulations, 43 J Regul Econ 248 (2013).

[81] See the catalogue in Ted Gayer and W. Kip Viscusi, Overriding Consumer Preferences With Energy Regulations, 43 J Regul Econ 248, 252 (2013).

Consumers can, of course, choose fuel-efficient vehicles if they like. If they do not, perhaps it is because the less fuel-efficient vehicles are smaller or less powerful, or because they are inferior along some other dimension. The behavioral question is whether consumers neglect fuel economy because of (for example) present bias or limited attention, or instead consider fuel economy but find it outweighed by other factors. The behavioral hunch is that present bias or limited attention do play a role, but a hunch is not evidence.[82] Behavioral welfare economics would carefully investigate the hunch, and consider consumer savings to the extent that the evidence suggests that they are real.[83] In fact that is a central, even defining question in contemporary regulatory policy, bearing on energy efficiency requirements as well as fuel-economy regulation.[84]

B. Defining Direct Judgments

What about people's direct judgments? Do they deserve deference? To make progress on this question, we have to know how to identify them, which means that we have to solve what might be called the level of abstraction problem. Return to the case of Norma. We could say that she prefers Yellow Box to Red Box, believing that Yellow Box contains the apple; that is clearly an indirect judgment, because she wants the apple, not the pear. But we could also say that what she wants is a good snack rather than a less good snack (bracketing the question of what, exactly, makes a snack good), and that the choice of an apple is an indirect judgment. Or we could say that what she wants is a good afternoon, or day, or year, or life, and the choice of a good snack is instrumental to one of those things.

Suppose that direct judgments are described at the highest level of abstraction – say, having a good life (on the assumption that a good life, properly defined, is intrinsically good). If so, essentially all judgments are indirect judgments, in the sense that they are meant as ways of getting a good life. The problem is that if direct judgments are described at the highest levels of abstraction, it is unclear what it means, if it means anything at all, to defer to individual judgment.

It would be possible to understand direct judgments at a lower level of abstraction. Choosers make judgments about what kind of day they want to have or what kind of life they want to lead. They might prefer high-calorie, full-sugar soda to diet drinks, or pizza to salad, even if they gain weight. They might like basketball but not football. They might want to devote themselves to family. They might want to devote themselves to some cause. They might want to pursue art or sport. They might want to marry, or not. They might want to live exciting, risk-filled lives, even at some cost to their health and longevity. Does anything in behavioral economics demonstrate that people are mistaken with respect to these judgments?

---

[82] The hunch is questioned in id. and in Hunt Allcott and Christopher Knittel, Are Consumers Poorly Informed About Fuel Economy? Evidence From Two Experiments, 11 Am. Econ J.: Economic Policy 1 (2019); it is supported in Kenneth Gillingham et al., Consumer Myopia in Vehicle Purchases (2019), available at https://www.nber.org/papers/w25845.

[83] The issue remains debated and in my view unresolved. See id.

[84] See Gayer and Viscusi, supra note; Allcott and Knittel, supra note.

That is not the easiest question to answer. It is one thing to insist that people choose inferior health care plans, given their situations,[85] that they make some bad food choices,[86] given their overall concerns, or that they do not purchase energy-efficient motor vehicles when it would be in their economic interest to do so.[87] It is quite another to insist that they choose the wrong sorts of days or lives. Do behavioral economists have anything to say on that question? If the answer is "yes," it might be for empirical reasons. A body of research in behavioral science points to "hedonic forecasting errors," which occur when people make mistaken predictions about the effects of outcomes or options on their subjective well-being.[88] Because the idea of "hedonic forecasting" is a bit narrow, we might ask whether people make "welfare forecasting errors," that is, whether they make mistaken judgments about what will increase their welfare. They might think that their lives will be better if they marry, but they might be wrong. If so, we might be tempted to say that their judgment was indirect, which would bring us back to the level of abstraction problem. And if they are making the wrong choices, it probably must, simply as a logical matter, be because of a lack of information or some behavioral bias, though we might not be able to identify it, and though it might not be part of the standard catalogue of behavioral biases.[89]

There is also the question of how to define welfare, such that we can identify a welfare forecasting error; I will get to that question momentarily. An inquiry into subjective well-being – undoubtedly a component of welfare – is more tractable. People might think that they would be happier in California, and therefore move there, but perhaps they would not be happier at all.[90] People might think that they would be unhappier if they left their current personal situation, but they might be quite mistaken on that point.[91]

Many philosophers embrace what they call "perfectionism," urging that some kinds of lives are simply better than others.[92] Some perfectionists are Aristotelian, emphasizing Aristotle's notion of functioning and an understanding of what it means for a human being to be

---

[85] See note supra.
[86] See Rabin, supra note.
[87] See Gillingham et al., supra note.
[88] Daniel Gilbert and Timothy Wilson, Miswanting: Some Problems in the Forecasting of Future Affective States, in *Feeling and Thinking: The Role of Affect in Social Cognition*, ed. Joseph Forgas (New York: Cambridge University Press, 2000), 186; Daniel Gilbert et al., Immune Neglect: A Source of Durability Bias in Affective Forecasting, *Journal of Personality and Social Psychology* 75, no. 3 (1998): 617–38.
[89] For a useful catalogue, see Cognitive Illusions (Rudiger F. Pohl ed. 2016).
[90] See Kahneman and Schkade, supra note. Bernheim and Dubinsky, supra note, raise a number of objections to reliance on subjective well-being. It is right to insist that subjective well-being is not the only thing that people care about; for example, choosers might sacrifice their subjective well-being for the sake of others, or for the sake of moral goals. But the case for relying on choices, rather than subjective well-being, is not made by that point. In some contexts, choices really are an effort to promote subjective well-being, and they go wrong.
[91] For evidence, see Steven Levitt, Heads or Tails: The Impact of A Coin Toss on Major Life Decisions and Subsequent Happiness (2016), available at https://www.nber.org/papers/w22487. In general, Levitt finds excessive caution in making life-changing decisions.
[92] See Thomas Hurka, Perfectionism (1996). For a defense of liberal perfectionism, see Joseph Raz, The Morality of Freedom (1985); for critiques of perfectionism, see John Rawls, Political Liberalism (1991); Sarah Conly, supra note. For a general account, see https://plato.stanford.edu/entries/perfectionism-moral/

truly human.[93] Some liberals are perfectionists; they give pride of place to a distinctive understanding of freedom.[94] Marx was a perfectionist.[95] Some forms of perfectionism are theological; consider Aquinas. On grounds of either autonomy (invoking Kant) or welfare (invoking Mill), we might want to reject perfectionism.[96] But to say the least, these are complicated normative questions, and nothing in standard economics, or behavioral economics, is likely to solve them.

My conclusions are straightforward. In some cases, people lack information.[97] In some cases, we can identify a behavioral market failure, in the sense that people fall prey to an identifiable behavioral bias, and their choices make their lives go worse by their own lights.[98] When this is so, some kind of corrective response is likely to be a good idea, perhaps in the form of a nudge, perhaps in the form of a tax, perhaps in the form of a mandate. In a free society, it nonetheless makes sense for those involved in law and policy to adopt a working presumption: *Each of us should be taken to be the best judge of what will promote our own well-being*, *to the extent that we are adequately informed and sufficiently free of behavioral biases.* In practice, that presumption can be disciplined by asking an assortment of subsidiary questions: What do informed people choose? What do active choosers choose? In circumstances in which people are free of (say) present bias, what do they choose? What do people choose when their viewscreen is broad and they do not suffer from limited attention?

In principle, that presumption allows considerable room for interference with indirect judgments; it authorizes means paternalism, often in the interest of increasing navigability. With respect to people's ends, operating at a high level of abstraction, those who offer the pragmatic presumption call for considerable deference to freedom of choice, on self-consciously pragmatic grounds. But in doing so, they recognize that the underlying justifications, founded on autonomy and welfare, are philosophical in nature. If they offer those justifications with conviction, they also do so with humility.

---

[93] See Philippa Foot, Natural Goodness (2001); Martha Nussbaum, Women and Human Development: The Capabilities Approach (2000), can be placed in the same extended family.
[94] See Raz, supra note.
[95] The best discussion remains Jon Elster, Making Sense of Marx (1985).
[96] An excellent, brisk discussion can be found in Conly, supra note, at 107-112.
[97] The point is emphasized in Oren Bar-Gill, Seduction by Contract (2012).
[98] See Bubb and Pildes, supra note, for a series of examples.