

THE YALE LAW JOURNAL

CASS R. SUNSTEIN

The Storrs Lectures: Behavioral Economics and Paternalism

ABSTRACT. A growing body of evidence demonstrates that in some contexts and for identifiable reasons, people make choices that are not in their interest, even when the stakes are high. Policymakers in a number of nations, including the United States and the United Kingdom, have used this evidence to inform regulatory initiatives and choice architecture. Both the resulting actions and the relevant findings have raised the possibility that an understanding of human errors opens greater space for paternalism (and thus raises doubts about John Stuart Mill's famous "harm principle"). Such errors can be thought of as behavioral market failures, and they are an important supplement to the standard account of market failures. Actions taken to correct behavioral market failures can sometimes be justified, even if the resulting actions are paternalistic. While hard forms of paternalism cannot be ruled out of bounds, a general principle of behaviorally informed regulation—its first and only law—is that the appropriate responses to behavioral market failures usually consist of nudges, generally in the form of disclosure, warnings, and default rules. Some people invoke autonomy as an objection to paternalism, but the strongest objections are welfarist in character. Official action may fail to respect heterogeneity, may diminish learning and self-help, may be subject to pressures from self-interested private groups (the problem of "behavioral public choice"), and may reflect the same errors that ordinary people make. Where paternalism is optional, the objections, though plausible, are unhelpfully abstract; they depend on empirical assumptions that may not hold in identifiable contexts. There are many opportunities for improving human welfare through improved choice architecture.

AUTHOR. Robert Walmsley University Professor and Felix Frankfurter Professor of Law, Harvard Law School. This Feature is a revised version of the Storrs Lectures in Jurisprudence, which were given at Yale Law School on November 12-13, 2012. A much-shortened version, intended for a general audience, will be included as a chapter in CASS R. SUNSTEIN, *SIMPLER: THE FUTURE OF GOVERNMENT* (forthcoming 2013). The author is grateful above all to Richard Thaler for comments and for many years of discussion of these topics. It is important to emphasize that Thaler does not agree with everything said here and hence that nothing here should be taken as reflective of a shared view about nudges and nudging. Special thanks to audiences at Yale Law School for graciousness and kindness, and for a host of valuable thoughts and suggestions. The author is also grateful to Esther Duflo, Elizabeth Emens, Christine Jolls, Martha Nussbaum, Eric Posner, Richard Posner, Lucia Reisch, and Adrian Vermeule for illuminating comments. The author is thankful as well to Daniel Kanter for excellent comments and research assistance and to participants in superb workshops at Harvard's Department of Economics, Harvard's Kennedy School of Government, and the University of Chicago Booth School of Business. From September 2009 to August 2012, the author served as Administrator of the White House Office of Information and Regulatory Affairs; nothing said here represents an official view in any way.



FEATURE CONTENTS

INTRODUCTION	1829
I. OCCASIONS FOR PATERNALISM?	1837
A. Two Systems in the Mind: Of Humans and Econs	1838
B. Behavioral Market Failures	1842
1. Present Bias and Time Inconsistency	1842
2. Ignoring Shrouded (but Important) Attributes	1845
3. Unrealistic Optimism	1848
4. Problems with Probability	1851
II. PATERNALISMS	1852
A. Working Definitions	1852
1. Choices and Welfare	1853
2. The (Important but Troubled) Distinction Between Means and Ends	1855
3. The (Important but Troubled) Distinction Between Hard and Soft	1858
4. A Very Quick Summary	1861
5. On Welfare	1861
B. The Paternalist's Large Toolbox	1864
III. AGAINST PATERNALISM: WELFARE	1867
A. Five Welfarist Objections: An Antipaternalist's Quintet	1868
B. Welfare: Normative Issues	1873
C. Welfare: Empirical Problems	1874
D. Imaginable Worlds and Rule-Consequentialist Antipaternalism	1877
E. Choice Architecture and Inevitable Nudges	1879
IV. AGAINST PATERNALISM: AUTONOMY	1881
A. Autonomy: The Thin Version	1881
B. Autonomy: The Thick Version	1882
C. Thin, Again	1883
D. Thick, Again	1885
E. An Accounting	1888
V. PROBLEMS WITH SOFT PATERNALISM: IMAGINARY AND REAL	1890

A. Of Transparency and Political Safeguards	1890
B. Of Easy Reversibility	1893
C. The Legitimate Claims of System 1	1894
D. A Real Concern: Impermissible Motivations	1897
CONCLUSION	1898

[T]he 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.

– John Stuart Mill¹

The central conundrum has been referred to as the Energy Paradox in this setting (and in several others). In short, the problem is that consumers appear not to purchase products that are in their economic self-interest. There are strong theoretical reasons why this might be so.

[1] Consumers might be myopic and hence undervalue the long-term;

[2] [Consumers] might lack information or a full appreciation of information even when it is presented;

[3] [Consumers] might be especially averse to the short-term losses associated with the higher prices of energy efficient products (the behavioral phenomenon of “loss aversion”);

[4] [E]ven if consumers have relevant knowledge, the benefits of energy efficient vehicles might not be sufficiently salient to them at the time of purchase

– U.S. Environmental Protection Agency²

INTRODUCTION

From 2009 to 2012, I was privileged to serve under President Obama as Administrator of the White House Office of Information and Regulatory Affairs, sometimes described (hyperbolically to be sure) as the nation’s “regulatory czar.” When I served in that position, the President stressed the need to consider flexible approaches that reduce costs and maintain freedom of choice for the American people. In fact, the President specifically charged me with promoting that goal, emphasizing its special importance in a period of serious economic difficulty.

In the Obama Administration, many of us were concerned about reducing

-
1. JOHN STUART MILL, ON LIBERTY 80 (David Bromwich & George Kateb eds., 2003) (1859).
 2. EPA, LIGHT-DUTY VEHICLE GREENHOUSE GAS EMISSION STANDARDS AND CORPORATE AVERAGE FUEL ECONOMY STANDARDS: EPA RESPONSE TO COMMENTS DOCUMENT FOR JOINT RULEMAKING 5-413 (2010), <http://www.epa.gov/oms/climate/regulations/420r10012a.pdf> (footnotes omitted).

regulatory costs, but we were alert both to the existence of standard market failures and to the findings of behavioral economics. We sought to identify approaches that would remedy those failures and respond to those findings, while surviving cost-benefit analysis and without imposing unjustified burdens on the private sector.³ We knew that there were many opportunities for using regulation to save both lives and money, by, among other things, promoting safety on the highways, cleaning the air, reducing smoking, increasing the fuel economy of cars, combating childhood obesity, and reducing health risks from food.

Executive Order 13,563—a document of signal importance, a kind of mini-constitution for the regulatory state—contains a key provision called “Flexible Approaches,” which states in no uncertain terms that “each agency shall identify and consider regulatory approaches that reduce burdens and maintain flexibility and freedom of choice for the public. These approaches include warnings, appropriate default rules, and disclosure requirements as well as provision of information to the public in a form that is clear and intelligible.”⁴ This provision can be understood as an explicit recognition of the potential value of low-cost, freedom-preserving approaches, or nudges.⁵

Behavioral economists have emphasized that in important contexts, people err.⁶ Human beings can be myopic and impulsive, giving undue weight to the short term (perhaps by smoking, perhaps by texting while driving, perhaps by eating too much chocolate).⁷ What is salient greatly matters.⁸ If an important

3. A catalogue of regulatory actions during the first term of the Obama Administration can be found in CASS R. SUNSTEIN, *SIMPLER: THE FUTURE OF GOVERNMENT* (forthcoming 2013) [hereinafter SUNSTEIN, SIMPLER]; and Cass R. Sunstein, *Empirically Informed Regulation*, 78 U. CHI. L. REV. 1349 (2011) [hereinafter Sunstein, *Empirically Informed*].
4. Exec. Order No. 13,563, 76 Fed. Reg. 3821 (Jan. 18, 2011), <http://www.gpo.gov/fdsys/pkg/FR-2011-01-21/pdf/2011-1385.pdf>.
5. See RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (2008).
6. An authoritative discussion is DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* (2011). On behavioral economics and public policy generally, see THALER & SUNSTEIN, *supra* note 5. In many respects, this Feature develops and attempts to deepen arguments produced jointly with Richard Thaler in *Nudge*, and for this reason, it may be useful to explain the relationship between this Feature and that book. First, this Feature, unlike the book, ventures a general treatment of the relationship between behavioral economics and paternalism; it discusses harder forms and is not limited to nudges. The effort to venture a more general treatment takes the analysis in a number of new directions. Second, this Feature draws on the author’s experience in the federal government, and a number of the examples, and many of the relevant concerns, come directly from that experience. Third, this Feature engages a number of illuminating recent discussions of behavioral economics and paternalism.
7. See David Laibson, *Golden Eggs and Hyperbolic Discounting*, 112 QJ. ECON. 443, 445

feature of a situation, an activity, or a product lacks salience, people might ignore it, possibly to their advantage (perhaps because it is in the other room, and fattening) and possibly to their detriment (if it could save them money or extend their lives). Human beings procrastinate and sometimes suffer as a result;⁹ they are greatly affected by default rules, potentially to their detriment.¹⁰ They can be unrealistically optimistic and for that reason make unfortunate and even dangerous choices.¹¹ People make “affective forecasting errors”: they predict that activities or products will have certain beneficial or adverse effects on their own well-being, but those predictions turn out to be wrong.¹²

It is important to emphasize that free markets provide significant protection against such errors. Most important, markets often deter exploitation of human fallibility. If companies provide unhelpful default rules, steering consumers in directions that harm them, they may be punished as a result of competition. Companies that shroud expensive attributes, costing consumers a lot of money, may find themselves without customers before long. In addition, companies offer countless services to help people counteract self-control problems. The market itself creates strong incentives for companies to respond to these and other behavioral problems. With new technologies, those responses will become increasingly helpful, frequent, inventive, and personalized;¹³ helpful “apps,” of countless sorts, are proliferating,¹⁴ and in the future, we will see unimaginably more. The market

(1997).

8. For a discussion of some of the foundational issues, see Pedro Bordalo, Nicola Gennaioli & Andrei Shleifer, *Salience in Experimental Tests of the Endowment Effect*, 102 AM. ECON. REV. 47 (2012); Pedro Bordalo, Nicola Gennaioli & Andrei Shleifer, *Salience Theory of Choice Under Risk*, 127 QJ. ECON. 1243 (2012).
9. See Ted O’Donoghue & Matthew Rabin, *Choice and Procrastination*, 116 QJ. ECON. 121, 121-22 (2001); Richard H. Thaler & Shlomo Benartzi, *Save More Tomorrow™: Using Behavioral Economics To Increase Employee Saving*, 112 J. POL. ECON. S164, S168-69 (2004).
10. See DUNCAN J. WATTS, EVERYTHING IS OBVIOUS ONCE YOU KNOW THE ANSWER: HOW COMMON SENSE FAILS US 30-31 (2011); Eric J. Johnson & Daniel G. Goldstein, *Decisions by Default*, in BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 417 (Eldar Shafir ed., 2013); Eric J. Johnson & Daniel Goldstein, *Do Defaults Save Lives?*, 302 SCIENCE 1338 (2003).
11. See TALIA SHAROT, THE OPTIMISM BIAS: A TOUR OF THE IRRATIONALLY POSITIVE BRAIN (2011).
12. See, e.g., Elizabeth W. Dunn, Daniel T. Gilbert & Timothy D. Wilson, *If Money Doesn’t Make You Happy, Then You Probably Aren’t Spending It Right*, 21 J. CONSUMER PSYCHOL. 115 (2011); Daniel T. Gilbert et al., *Immune Neglect: A Source of Durability Bias in Affective Forecasting*, 75 J. PERSONALITY & SOC. PSYCHOL. 617 (1998).
13. See Richard H. Thaler & Will Tucker, *Smarter Information, Smarter Consumers*, HARV. BUS. REV., Jan.-Feb. 2013, <http://hbr.org/2013/01/smarter-information-smarter-consumers/ar/1>.
14. See, e.g., APPS FOR HEALTHY KIDS, <http://appsforhealthykids.com> (last visited Feb. 21, 2013); Jocelyn K. Gleib, *10 Online Tools for Better Attention & Focus*, 99U, <http://99u.com/articles/6969/10-Online-Tools-for-Better-Attention-Focus> (last visited Feb. 21, 2013); Paige

for protecting people against their own mistakes is flourishing.¹⁵

But there is another side. In free markets, some sellers attempt to exploit human errors, and the forces of competition may turn out to reward, rather than punish, such exploitation. In identifiable cases, those who do *not* exploit human errors will be seriously punished by market forces, simply because their competitors are doing so and profiting as a result. Credit markets provide many examples in the domains of cell phones, credit cards, and mortgages.¹⁶ More generally, some policies will not be designed well if they are not informed by what we know about human behavior.¹⁷

It is true, of course, that a great deal remains to be understood about the nature of human error in disparate contexts. Research is continuing, and more is being learned every day; some behavioral findings are highly preliminary and need further testing. There is much that we do not know. Randomized controlled trials, the gold standard for empirical research, must be used much more to obtain a better understanding of how the relevant findings operate in the world.¹⁸ Even at this stage, however, the underlying findings have been widely noticed, and behavioral economics and related fields have had a significant effect on policies in several nations, including the United States and the United Kingdom.

In the United States, a number of initiatives have been informed by relevant empirical findings, and behavioral economics has played an unmistakable role in numerous domains. These initiatives enlist tools such as disclosures, warnings, and default rules, and they can be found in multiple areas, including fuel economy, energy efficiency, environmental protection, health care, and obesity.¹⁹ As a result, behavioral findings have become an important reference point for regulatory and other policymaking in the United

Wachner, *Best Exercise and Fitness Apps for iPhone and iPod Touch: Let Your iPod Be Your Trainer*, ABOUT.COM, <http://exercise.about.com/od/videosmusicsoftware/tp/fitnessapps.htm> (last updated Oct. 30, 2012).

15. For valuable discussion, see IAN AYRES, *CARROTS AND STICKS* (2010).
16. See OREN BAR-GILL, *SEDUCTION BY CONTRACT: LAW, ECONOMICS, AND PSYCHOLOGY IN CONSUMER MARKETS* (2012).
17. See Andrei Shleifer, *Psychologists at the Gate: A Review of Daniel Kahneman's Thinking, Fast and Slow*, 50 *J. ECON. LITERATURE* 1080 (2012).
18. Michael Greenstone, *Toward a Culture of Persistent Regulatory Experimentation and Evaluation*, in *NEW PERSPECTIVES ON REGULATION* 111 (David Moss & John Cisternino eds., 2009). For a number of discussions of randomized controlled trials, including nudges, see ABHIJIT V. BANERJEE & ESTHER DUFLO, *POOR ECONOMICS: A RADICAL RETHINKING OF THE WAY TO FIGHT GLOBAL POVERTY* (2011).
19. See Sunstein, *Empirically Informed*, *supra* note 3.

States.²⁰

In the United Kingdom, Prime Minister Cameron has created a Behavioural Insights Team with the specific goal of incorporating an understanding of human behavior into policy initiatives.²¹ The official website states that its “work draws on insights from the growing body of academic research in the fields of behavioural economics and psychology which show how often subtle changes to the way in which decisions are framed can have big impacts on how people respond to them.”²² The team has used these insights to promote initiatives in numerous areas, including smoking cessation, energy efficiency, organ donation, consumer protection, and compliance strategies in general.²³ Other nations have expressed interest in the work of the team, and its operations are expanding.²⁴

Behavioral economics has drawn attention in Europe more broadly. The Organisation for Economic Cooperation and Development has published a Consumer Policy Toolkit that recommends a number of initiatives rooted in behavioral findings.²⁵ In the European Union, the Directorate-General for Health and Consumers has also shown the influence of behavioral economics.²⁶ A report from the European Commission, called *Green Behavior*, enlists behavioral economics to outline policy initiatives to protect the environment.²⁷

These developments, and the relevant findings, raise a natural question, which is whether an understanding of human behavior opens greater space for paternalism, supplementing the standard accounts of market failures by

20. See, e.g., *id.*; see also Theresa M. Marteau, Gareth J. Hollands & Paul C. Fletcher, *Changing Human Behavior To Prevent Disease: The Importance of Targeting Automatic Processes*, 337 SCIENCE 1492 (2012) (exploring the role of automatic processing in behavior in the domain of health).

21. See *Behavioural Insights Team*, CABINET OFF., <http://www.cabinetoffice.gov.uk/behavioural-insights-team> (last visited Jan. 28, 2013).

22. *Id.*

23. Various reports can be found on the website of the Behavioural Insights Team. See *id.*

24. See Oliver Wright, *Steve Hilton’s ‘Nudge Unit’ Goes Global*, INDEPENDENT (London), Sept. 20, 2012, <http://www.independent.co.uk/news/uk/politics/steve-hiltons-nudge-unit-goes-global-8157492.html>.

25. See *Consumer Policy Toolkit*, ORG. FOR ECON. COOPERATION & DEV. (July 2010), <http://www.oecd.org/sti/consumerpolicy/consumerpolicytoolkit.htm>.

26. See Directorate-Gen. for Health & Consumers, *Consumer Behaviour: The Road to Effective Policy-Making*, EUR. COMMISSION (2010), <http://ec.europa.eu/consumers/docs/1dg-sanco-brochure-consumer-behaviour-final.pdf>.

27. See Sci. for Env’t Policy, *Future Brief: Green Behavior*, EUR. COMMISSION (Oct. 2012), <http://ec.europa.eu/environment/integration/research/newsalert/pdf/FB4.pdf>. A number of relevant sources can be found at *Resources*, INUDGEYOU, <http://www.inudgeyou.com/resources> (last visited Jan. 28, 2013).

providing grounds for government action even in the absence of harm to others or some kind of collective action problem.²⁸ We know, for example, that people are greatly affected by *choice architecture*, understood as the social background against which choices are made.²⁹ Such architecture is pervasive and inevitable, and it greatly influences outcomes. In fact, it can be decisive. It effectively makes countless decisions for us, or at least affects our decisions.³⁰ Choice architecture exists whenever we enter a cafeteria, a restaurant, a hospital, or a grocery store; when we select a mortgage, a car, a health care plan, or a credit card; when we turn on a tablet or a computer and visit our favorite websites; and when we apply for drivers' licenses or building permits or social security benefits. For all of us, a key question is whether the relevant choice architecture is helpful and simple or harmful, complex, and exploitative.

Should choice architects, including those in the public sphere, be authorized to move people's decisions in their preferred directions? Would any such efforts be unacceptably paternalistic? Who will monitor the choice architects, or create a choice architecture for them?³¹ From various empirical findings, it is possible to identify a set of *behavioral market failures*,³² understood as a set of market failures that complement the standard economic account and that stem from human error. Is it unacceptably paternalistic to use such failures to justify regulation, even when externalities are not involved? Is it legitimate to use choice architecture to counteract behavioral market failures?

My goal here is to explore these questions. My basic answer is that behavioral market failures do, in fact, justify paternalism.³³ When such failures

28. A positive answer is provided in SARAH CONLY, *AGAINST AUTONOMY: JUSTIFYING COERCIVE PATERNALISM* (2013). A negative answer is provided in RICCARDO REBONATO, *TAKING LIBERTIES: A CRITICAL EXAMINATION OF LIBERTARIAN PATERNALISM* (2012); Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133 (2006); and Joshua D. Wright & Douglas H. Ginsburg, *Behavioral Law and Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 NW. U. L. REV. 1033 (2012).

29. See THALER & SUNSTEIN, *supra* note 5, at 81-100.

30. For one example, see Paul Rozin et al., *Nudge to Nobesity I: Minor Changes in Accessibility Decrease Food Intake*, 6 JUDGMENT & DECISION MAKING 323, 329 (2011).

31. It should be emphasized, however, that many behaviorally informed approaches, such as the simplification of complex requirements, need not have a paternalistic dimension. On choice architecture for choice architects, see SUNSTEIN, *SIMPLER*, *supra* note 3.

32. See BAR-GILL, *supra* note 16, at 2-4.

33. My emphasis here is on behavioral market failures as a supplement to standard market failures. It is true, of course, that there are other justifications for government action, falling in neither category. We might believe, for example, that prohibitions on discrimination of various kinds, or protections of privacy, are justified even if there is no behavioral or standard market failure. For one catalogue, see CASS R. SUNSTEIN, *AFTER THE RIGHTS REVOLUTION: RECONCEIVING THE REGULATORY STATE* 47-73 (1990).

exist, and are significant, there are good (presumptive) reasons for a regulatory response even when no harm to others can be found. But because of heterogeneity and the risk of government error, it is usually best to use the mildest and most choice-preserving forms of intervention, such as nudges.³⁴ We might even venture a general principle, which might be called the first (and only) law of behaviorally informed regulation: *in the face of behavioral market failures, the best responses usually are disclosures of information, warnings, default rules, and other kinds of nudges, at least when there is no harm to others.* But there are exceptions to the general principle, and the choice of response depends on an analysis of costs and benefits.³⁵ In some cases, no response at all may be best, because the costs exceed the benefits. In other cases, stronger responses, even mandates, may turn out to be justified, because the benefits exceed the costs. Social welfare is the master concept, and when social welfare calls for a stronger response, we should give it serious consideration.³⁶

It is useful to begin, I suggest, by distinguishing among varieties of paternalism. Some varieties respect people's ends and try only to influence their choice of means; other varieties attempt to affect people's choices of ends. *Means paternalists* might encourage (or perhaps even require) people to save money with refrigerators that are inexpensive to operate, when saving money is exactly what they want. *Ends paternalists* might forbid people from engaging in certain sexual activity, even though engaging in such activity is exactly what they want. Behavioral economists generally favor paternalism about means, not ends. Most of their key findings involve human errors with respect to means; their goal is to create choice architecture that will make it more likely that people will promote their own ends.

Moreover, some varieties of paternalism are highly aggressive, or "hard," while others are weaker, or "soft." Soft paternalism is libertarian, in the sense that it preserves freedom of choice.³⁷ A jail sentence and a criminal fine count as

34. See THALER & SUNSTEIN, *supra* note 5, at 8.

35. Consider, for example, the domains of energy efficiency and fuel economy, where an analysis of costs and benefits may support stronger forms of paternalism. On the general topic of coercion, emphasizing the role of costs and benefits, see CONLY, *supra* note 28, which argues in favor of coercive paternalism. Notably, the interest in nudging and in soft paternalism has been controversial among those who emphasize that mandates and bans may be necessary. See, e.g., *id.*; George Loewenstein & Peter O'Donoghue, *Economics Behaving Badly*, N.Y. TIMES, July 14, 2010, <http://www.nytimes.com/2010/07/15/opinion/15loewenstein.html>.

36. On some of the foundational issues, see MATTHEW D. ADLER, *WELL-BEING AND FAIR DISTRIBUTION: BEYOND COST-BENEFIT ANALYSIS* (2011); with respect to paternalism, see CONLY, *supra* note 28, at 7, which suggests that whether the benefits justify the costs "is the only determinant of acceptability."

37. See THALER & SUNSTEIN, *supra* note 5, at 4-6.

hard paternalism, whereas a disclosure policy, a warning, and a default rule count as soft or libertarian paternalism. Some forms of paternalism impose material costs on people's choices in order to improve their welfare; other forms of paternalism impose affective or psychic costs. Behavioral economists have generally favored soft rather than hard paternalism.³⁸ Means paternalism can be hard or soft, and the same is true of ends paternalism. My topic here extends far beyond libertarian paternalism and nudges, understood as approaches that affect choices without coercion, but it is important to see that nudges generally fall in the categories of means paternalism and soft paternalism.³⁹

My central claim is simple: behavioral market failures are an important supplement to the standard account of market failures, and in principle they do justify (ideal) responses, even if those responses are paternalistic. As in the case of standard failures, however, the argument for a government response must be qualified by a recognition that the cure may be worse than the disease, and that all relevant benefits and costs must be taken into account.

I offer four additional conclusions:

1. Choice architecture is inevitable, and hence certain influences on choices are also inevitable, whether or not they are intentional or a product of any kind of conscious design.
2. Some of the most intuitively appealing objections to paternalism rely on autonomy, but as applied to most efforts to remedy behavioral market failures, those objections lack force, because such efforts do not interfere with autonomy, rightly understood. In fact, some such efforts promote autonomy, in part because they open up time and resources for more pressing matters.⁴⁰ There is also a risk that some of these autonomy-based objections are rooted in a heuristic for what really matters, which is welfare.
3. The most powerful objections to paternalism are welfarist in character. In many contexts, those objections are a good place to start and possibly to end, especially insofar as they emphasize the importance of private learning and the risk of government error. But they depend on normative claims that are complex and highly

38. See, e.g., *id.*; see also CONLY, *supra* note 28, at 12 (suggesting that "paternalistic regulations are designed to help us reach our own goals.").

39. See THALER & SUNSTEIN, *supra* note 5, at 6-8.

40. See Anuj K. Shah, Sendhil Mullainathan & Eldar Shafir, *Some Consequences of Having Too Little*, 338 SCIENCE 682 (2012).

contested, and on empirical claims that are often false. There is no sufficient abstract or a priori argument against paternalism, whether hard or soft.⁴¹

4. The welfarist arguments against paternalism, new or old, are irrelevant insofar as choice architecture, and nudges, are inevitable. But insofar as paternalism is optional (and it often is), there is an intelligible rule-consequentialist objection to paternalism—though the strength of the argument depends on the form of paternalism. There are plausible rule-consequentialist arguments against (optional) paternalism, but those arguments depend on strong empirical assumptions, involving extreme optimism about markets and extreme pessimism about public officials, that are unlikely to hold in our world. The objections to paternalism are weakest when it is soft and limited to means; especially in such cases, there are many opportunities for improving welfare without intruding on freedom of choice.

The remainder of the discussion comes in five parts. Part I discusses human errors, with particular emphasis on those errors that are most likely to matter for purposes of regulatory policy. Part II explores the nature of paternalism, distinguishing among various forms, and emphasizing the wide range of tools that paternalistic choice architects might use. Part III turns to welfarist objections to paternalism. Part IV explores autonomy. Part V discusses several independent objections to soft or libertarian paternalism, particularly those that emphasize the potential lack of transparency, the risk of manipulation, and the limits of reversibility. The discussion ends with a brief conclusion.

I. OCCASIONS FOR PATERNALISM?

In recent decades, there has been an outpouring of empirical work on human cognition and the risk of error.⁴² As noted, this work has been noticed

41. An especially valuable treatment of these issues is Esther Duflo, Abdul Latif Jameel Professor of Poverty Alleviation & Dev. Econ., Mass. Inst. of Tech., Tanner Lectures on Human Values and the Design of the Fight Against Poverty (May 2, 2012), <http://economics.mit.edu/files/7904>.

42. See generally *ADVANCES IN BEHAVIORAL ECONOMICS* (Colin F. Camerer et al. eds., 2003); *2 ADVANCES IN BEHAVIORAL FINANCE* (Richard H. Thaler ed., 2005); *CHOICES, VALUES, AND FRAMES* (Daniel Kahneman & Amos Tversky eds., 2000); *HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT* (Thomas Gilovich et al. eds., 2002).

by policymakers,⁴³ and its influence is likely to grow in coming decades. My goal here is to provide a brief summary, acknowledging that research is continuing and that a great deal remains to be learned, and emphasizing those findings that have special importance for exploring regulation and the question of paternalism.

A. Two Systems in the Mind: Of Humans and Econs

Within recent social science, authoritatively discussed by Daniel Kahneman in his masterful *Thinking, Fast and Slow*, it has become standard to suggest that the human mind contains not one but two “cognitive systems.”⁴⁴ In the social science literature, the two systems are unimaginatively described as System 1 and System 2.⁴⁵ System 1 is the automatic system, while System 2 is more deliberative and reflective. System 1 can be understood to reflect the behavior of Humans, whereas Econs think and act in accordance with System 2.⁴⁶

System 1 works fast. Much of the time, it is on automatic pilot. It is driven by habits. It can be emotional and intuitive. When it hears a loud noise, it is inclined to run. When it is offended, it wants to hit back. It certainly eats a delicious brownie. It can procrastinate; it can be impulsive. It wants what it wants when it wants it. It can be excessively fearful and too complacent. It is a doer, not a planner. System 1 is a bit like Homer Simpson, James Dean (from *Rebel Without a Cause*), and Pippi Longstocking.

System 2 is more like a computer or Mr. Spock from the old *Star Trek* show (or the android Data from the somewhat-less-old *Star Trek* show). It is deliberative. It calculates. It hears a loud noise, and it assesses whether the noise is a cause for concern. It thinks about probability, carefully though sometimes slowly. It does not really get offended. If it sees reasons for offense, it makes a careful assessment of what, all things considered, ought to be done. It sees a delicious brownie, and it makes a judgment about whether, all things considered, it should eat it. It insists on the importance of self-control. It is a planner more than a doer.

At this point, it might be asked: What, exactly, are these systems? Are Humans and Econs agents? Do they operate as homunculi in the brain? Are they little people? Are they actually separate? In the case of conflict, who adjudicates? The best answer is that the idea of two systems is a heuristic

43. See Sunstein, *Empirically Informed*, *supra* note 3, at 1370-78.

44. KAHNEMAN, *supra* note 6; see also THALER & SUNSTEIN, *supra* note 5, at 6-8 (discussing “Humans” and “Econs”).

45. See KAHNEMAN, *supra* note 6, at 13.

46. See THALER & SUNSTEIN, *supra* note 5, at 19-22.

device, a simplification that is designed to refer to automatic, effortless processing and more complex, effortful processing. When people are asked to add one plus one, or to walk from their bedroom to their bathroom in the dark, or to read the emotion on the face of their best friend, the mental operation is easy and rapid. When people are asked to multiply 179 by 283, or to navigate a new neighborhood by car, or to decide which retirement or health insurance plan best fits their needs, the mental operation is difficult and slow.

Identifiable regions of the brain are active in different tasks, and hence it may well be right to suggest that the idea of “systems” has physical referents. An influential discussion states that “[a]utomatic and controlled processes can be roughly distinguished by where they occur in the brain.”⁴⁷ The prefrontal cortex, the most advanced part of the brain (in terms of evolution) and the part that most separates human beings from other species, is associated with deliberation and hence with System 2. The amygdala has been associated with a number of automatic processes, including fear,⁴⁸ and can thus be associated with System 1.

With respect to intertemporal choice (an especially important topic for behaviorally informed regulation), it has been found that when impatient people are thinking about their future selves, the particular region of the brain that is most active when people are thinking about themselves is significantly less active.⁴⁹ In patient people, by contrast, that region of the brain is significantly more active when they are thinking of their future selves.⁵⁰ Here, then, is a neurological basis for distinguishing not only between Humans and Econs but also between different members of the human species. This finding has clear implications for myopia, in the form of neglect of the future, and time inconsistency.⁵¹ In neural terms, impatient people think of their future selves in the same way that they think of strangers – raising the possibility that they may not be sufficiently concerned about their own future well-being.⁵² Neural evidence also suggests that when people’s emotions are strongly engaged, in a way that makes them motivated to accept certain political conclusions, identifiable features of the brain are active – and that when people do not have

47. Colin Camerer et al., *Neuroeconomics: How Neuroscience Can Inform Economics*, 43 J. ECON. LITERATURE 9, 17 (2005).

48. See Elizabeth A. Phelps, *The Human Amygdala and the Control of Fear*, in *THE HUMAN AMYGDALA* 204 (Paul J. Whalen & Elizabeth A. Phelps eds., 2009).

49. See Jason P. Mitchell et al., *Medial Prefrontal Cortex Predicts Intertemporal Choice*, 23 J. COGNITIVE NEUROSCIENCE 1, 6 (2010).

50. *Id.*

51. See *infra* Subsection I.B.1.

52. Mitchell et al., *supra* note 49, at 5.

a significant emotional stake, those regions are relatively inactive.⁵³

On the other hand, different parts of the brain interact, and it is not necessary to make technical or controversial claims about neuroscience in order to distinguish between effortless and effortful processing. The idea of System 1 and System 2 is designed to capture that distinction in a way that works for purposes of exposition (and that can be grasped fairly immediately by System 1).

Here is a striking demonstration of the relationship between System 1 and System 2: some of the most important cognitive errors, including several of relevance here (framing and loss aversion), *disappear when people are using a foreign language*.⁵⁴ Asked to resolve problems in a language that is not their own, people are less likely to blunder. In an unfamiliar language, they are more likely to get the right answer. How can this be?

The answer is straightforward. When people are using their own language, they think quickly and effortlessly, so System 1 has the upper hand. When people are using another tongue, System 1 is a bit overwhelmed, and may even be rendered inoperative, while System 2 is given a serious boost. Our rapid, intuitive reactions are slowed down when we are using a language with which we are not entirely familiar. We are more likely to do some calculating and to think deliberately—and at least on some questions, to give the right answers.⁵⁵ In a foreign language, people have some distance from their intuitions, and that distance can stand them in good stead. In a foreign language, Humans recede in favor of Econs.

There is a lesson here about the importance of technocratic approaches to law and regulation, including those that emphasize the need for careful consideration of costs and benefits.⁵⁶ Such approaches do not (exactly) use a foreign language, but they do ensure a degree of distance from people's initial judgments, thus constraining the mistakes associated with System 1. People do not naturally think about risk regulation in terms of costs and benefits, but the effort to do so can weaken or eliminate the effect of intuitions, in a way that

53. See Drew Westen et al., *Neural Bases of Motivated Reasoning: An fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election*, 18 J. COGNITIVE NEUROSCIENCE 1947, 1955 (2006).

54. Boaz Keysar, Sayuri L. Hayakawa & Sun Gyu An, *The Foreign-Language Effect: Thinking in a Foreign Tongue Reduces Decision Biases*, 23 PSYCHOL. SCI. 661 (2012).

55. Compare Samuel Beckett's decision to write some of his greatest works in French rather than English. Beckett said that the French language had an "aura of unfamiliarity about it," and that it allowed him to "escape the habits inherent in the use of a native language." THE GROVE COMPANION TO SAMUEL BECKETT: A READER'S GUIDE TO HIS WORKS, LIFE, AND THOUGHT 206 (C.J. Ackerly & S.E. Gontarski eds., 2004).

56. See, e.g., W. KIP VISCUSI, RATIONAL RISK POLICY (1998).

leads to greatly improved decisions.⁵⁷ There is also a point here about the hazards of relying on intuitions as a foundation for political or moral theory—a point to which I will return.⁵⁸

The defining feature of System 1 is that it is automatic, but I have said that System 1 can be emotional, and when it is, its emotional character creates both risks and opportunities. People may be immediately fearful of some risk—say, the risk associated with terrorism, or the risk of losses in the stock market—whether or not reality, and the relevant statistics, suggest that there is cause for alarm. A great deal of work finds that people tend to assess products, activities, and other people through “an affect heuristic.”⁵⁹ When the affect heuristic is at work, people evaluate benefits, costs, and probabilities not by running the numbers, but by consulting their feelings. They might hate coal-fired power plants or love renewable fuels, and those feelings may influence their judgments about the benefits and costs of coal-fired power plants and renewable fuels.⁶⁰ System 1 is doing the key work here.

In fact, some goods and activities come with an “affective tax” or an “affective subsidy,” in the sense that people like them more, or less, because of the affect that accompanies them. Advertisers, and public officials, try to create affective taxes and subsidies; consider public educational campaigns designed to reduce smoking or texting while driving. Some political campaigns have the same goal, attempting to impose a kind of affective tax on the opponent, and to enlist the affect heuristic in their favor. Many political campaigns appeal directly to System 1, not System 2. The same is true for some lawyers involved in trials or even appellate litigation. If System 1 can be enlisted, it may run the show, with System 2 operating as a kind of *ex post* helper.⁶¹ In many cases, System 2 acts as lawyer for the cause, and System 1 is a most demanding client.⁶²

One explanation for the operation of heuristics is that people decline to answer a hard question and answer a simpler one instead.⁶³ For political

57. For a detailed discussion of this subject, see SUNSTEIN, SIMPLER, *supra* note 3; and VISCUSI, *supra* note 56.

58. See *infra* Section IV.D.

59. See PAUL SLOVIC, *THE FEELING OF RISK: NEW PERSPECTIVES ON RISK PERCEPTION* 3-20 (2010).

60. See *id.*

61. See JONATHAN HAIDT, *THE RIGHTEOUS MIND: WHY GOOD PEOPLE ARE DIVIDED BY POLITICS AND RELIGION* (2012) (arguing that many judgments are automatic and intuitive, and explaining how our deliberative system works hard to justify those judgments *ex post*).

62. See *id.*

63. See KAHNEMAN, *supra* note 6, at 97-99.

candidates, people might not ask, “Do I agree with Candidate *A* or Candidate *B* on economic policy?” (a potentially complex question) but instead, “Do I like and trust this person?” or, “Is this person like me?” (potentially much easier questions).⁶⁴ Something similar is at work in educational campaigns that attempt to trigger fear (in the context, for example, of smoking, obesity, and texting while driving), and thus to engage System 1 rather than to offer statistical analyses.

B. Behavioral Market Failures

I now turn to four sets of mistakes that can lead to significant harms and that should be counted as behavioral market failures. As we shall see, all of these mistakes are firmly rooted in the operations of System 1. The unifying theme is that insofar as people are making the relevant errors, their choices will not promote their own ends. It follows that a successful effort to correct these errors would generally substitute an official judgment for that of choosers only with respect to means, not ends.⁶⁵ There are, however, some complexities in this claim. The distinction between means and ends raises a number of difficult puzzles, some of them involving the identification of people’s ends over time.

1. Present Bias and Time Inconsistency

According to standard economic theory, people will consider both the short term and the long term. They will take account of relevant uncertainties; the future is unpredictable, and significant changes may occur over time. People will appropriately discount the future. It is probably far better to have money, or a good event, a week from now than a decade from now. People may, rationally and reasonably, select different balances between the present and the future. With respect to present and future consumption, people who are twenty-five make different tradeoffs from people who are sixty-five, and for excellent reasons.

In practice, however, some people procrastinate or neglect to take steps that

64. For a related finding in the political domain, testifying to the power of System 1, see Christopher H. Achen & Larry M. Bartels, *Blind Retrospection: Electoral Responses to Drought, Flu, and Shark Attacks* (Fundación Juan March, Working Paper No. 2004/199, 2004), http://www.march.es/ceacs/publicaciones/working/archivos/2004_199.pdf.

65. See, e.g., Hunt Allcott, Sendhil Mullainathan & Dmitry Taubinsky, *Energy Policy with Externalities and Internalities* (Nat’l Bureau of Econ. Research, Working Paper No. 17,977, 2012), <http://www.nber.org/papers/w17977.pdf> (explaining that people often do not consider the long-term effects of energy efficiency and exploring how policies might help them to do so, thus combating “internalities”).

impose small, short-term costs but produce large, long-term gains, and at least some of the relevant actions seem hard to justify.⁶⁶ While System 2 considers the long term, System 1 is myopic, and, in multiple ways, people show *present bias*.⁶⁷ People may, for example, delay enrolling in a retirement plan,⁶⁸ starting to exercise, ceasing to smoke, or using some valuable, cost-saving technology.⁶⁹ In many cases, inertia is an exceedingly powerful force.⁷⁰

One implication is that some people fail to make choices that have short-term net costs but long-term net benefits—as is the case, for some, with choosing more energy-efficient products, including appliances and cars with good fuel economy.⁷¹ Another implication is that some people make choices that have short-term net benefits but long-term net costs, including a significant risk of causing premature death (as is the case, for many, with smoking cigarettes). Procrastination, inertia, hyperbolic discounting,⁷² and associated problems of self-control⁷³ are especially troublesome when the result is a small short-term gain at the expense of large long-term losses. There is a

66. See O'Donoghue & Rabin, *supra* note 9, at 121–22; Thaler & Benartzi, *supra* note 9, at S168–69. In the context of poverty, see BANERJEE & DUFLO, *supra* note 18, at 64–68. For an important and relevant discussion that does not involve procrastination but cognitive load, see Shah et al., *supra* note 40.
67. See, e.g., Jess Benhabib, Alberto Bisin & Andrew Schotter, *Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs*, 69 GAMES & ECON. BEHAV. 205 (2010).
68. Cf. Dean Karlan et al., *Getting to the Top of Mind: How Reminders Increase Saving* 1, 14 (Yale Econ. Dep't, Working Paper No. 82, 2010), <http://www.econ.yale.edu/ddp/ddp75/ddp0082.pdf> (showing the value of reminders in getting people to attend to savings).
69. See Esther Duflo, Michael Kremer & Jonathan Robinson, *Nudging Farmers To Use Fertilizer: Evidence from Kenya*, 101 AM. ECON. REV. 2350, 2351–54 (2011) (finding that farmers in western Kenya do not make economically advantageous fertilizer investments, but that a small, time-limited discount on the cost of fertilizer can increase investments, thus producing higher welfare than either a laissez-faire approach or large subsidies).
70. Inertia also helps account for the power of default rules. See Johnson & Goldstein, *supra* note 10; Cass R. Sunstein, *Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych* (SSRN Elec. Library, Working Paper No. 2,171,343, 2012), <http://ssrn.com/abstract=2171343>. An important discussion, with many implications, is Raj Chetty et al., *Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark* (Nat'l Bureau of Econ. Research, Working Paper No. 18,565, 2012), <http://www.nber.org/papers/w18565>.
71. See Hunt Allcott & Nathan Wozny, *Gasoline Prices, Fuel Economy, and the Energy Paradox* (Nat'l Bureau of Econ. Research, Working Paper No. 18,583, 2012), <http://www.nber.org/papers/w18583>.
72. See Laibson, *supra* note 7, at 445.
73. See Richard H. Thaler & H.M. Shefrin, *An Economic Theory of Self-Control*, 89 J. POL. ECON. 392 (1981). For an interesting application, discussed in more detail below, see Jonathan H. Gruber & Sendhil Mullainathan, *Do Cigarette Taxes Make Smokers Happier?*, 5 ADVANCES ECON. ANALYSIS & POL'Y 1, 20–21 (2005).

close connection between procrastination and myopia, understood as an excessive focus on the short term.⁷⁴

The problem of time inconsistency arises when people's preferences at Time One diverge from their preferences at Time Two.⁷⁵ At Time One, people might prefer to eat a great deal, to smoke, to spend, to become angry, to drink, to procrastinate, or to gamble. The resulting choices might have serious adverse effects on the same people at Time Two, leading to a significant welfare loss. As I have suggested, an identifiable region of the brain is most actively engaged when people are thinking about themselves, and for impatient people in particular, this region is less active when they are thinking about their future selves.⁷⁶ Studying the ventromedial prefrontal cortex (vMPFC), psychologist Jason Mitchell and his coauthors state that this "neural signature" suggests that "shortsighted decision-making occurs in part because people fail to consider their future interests as belonging to the self."⁷⁷ Thus, for those who are shortsighted, the "vMPFC response was nearly identical when people tried to predict their future enjoyment . . . and another person's present enjoyment," suggesting that such people think of their own future selves in the same way that they think of strangers.⁷⁸

Strikingly, Mitchell and his coauthors find the following:

[T]he magnitude of this vMPFC difference between judgments of present and future enjoyment predicted the impatience or shortsightedness of people's intertemporal choices. Those participants in whom vMPFC activity *most* differentiated between predictions of present and future enjoyment tended to make the most impatient decisions, preferring small present rewards to large future rewards. In contrast, participants in whom vMPFC did not differentiate between predictions of present and future enjoyment tended to make the most patient decisions, preferring large future rewards to small present rewards.⁷⁹

74. See generally Shlomo Benartzi & Richard H. Thaler, *Myopic Loss Aversion and the Equity Premium Puzzle*, 110 QJ. ECON. 73, 88 (1995).

75. See Haiyan Shui & Lawrence M. Ausubel, *Time Inconsistency in the Credit Card Market* (SSRN Elec. Library, Working Paper No. 586,622, 2004), <http://ssrn.com/abstract=586622>. For a technical treatment, see Roland Bénabou & Marek Pycia, *Dynamic Inconsistency and Self-Control: A Planner-Doer Interpretation*, 77 ECON. LETTERS 419 (2002).

76. See Mitchell et al., *supra* note 49, at 4-5.

77. *Id.* at 1.

78. *Id.* at 5.

79. *Id.* at 6.

Some behavioral economists have emphasized the problem of “internalities”⁸⁰—problems of self-control and errors in judgment that harm the people who make those very judgments. We can think of internalities as occurring when we make choices that injure our future selves. Of course people can use various techniques to overcome this problem, including precommitment strategies; consider Ulysses and the Sirens.⁸¹ As I have noted, private markets are perfectly capable of creating products and practices to help overcome self-control problems; in fact there are countless such products and practices. But it is at least plausible to suggest that regulatory approaches that address internalities can produce large welfare gains, in some cases by saving lives.⁸²

Such approaches might take the form of disclosure requirements or warnings, designed to promote self-control. Flexible approaches of this kind have the advantage of maintaining freedom of choice and thus respecting heterogeneity, which is especially important in light of the fact that reasonable people can trade off the present and the future in multiple ways based on the particulars of their situation. But in imaginable cases, an economic incentive or a mandate might be the best solution; consider, for example, efforts to promote healthy foods or bans on texting while driving, if understood to protect drivers (as well as those whom they endanger). With respect to internalities, energy policy includes many examples, such as energy-efficiency requirements for appliances and fuel-economy requirements for vehicles.⁸³ Under imaginable assumptions about costs and benefits, the best approach to a palpable neglect of the long term might turn out to be a ban.⁸⁴

2. Ignoring Shrouded (but Important) Attributes

What do people notice? What do they miss? In the late 1990s, social scientists Christopher Chabris and Daniel Simons tried to make some progress on these questions by asking people to watch a ninety-second movie, in which six ordinary people pass a basketball to one another.⁸⁵ The simple task? To

80. See, e.g., Allcott et al., *supra* note 65; R.J. Herrnstein et al., *Utility Maximization and Melioration: Internalities in Individual Choice*, 6 J. BEHAV. DECISION MAKING 149 (1993); Jonathan Gruber, *Smoking's 'Internalities'*, REGULATION, Winter 2002-2003, at 52.

81. See AYRES, *supra* note 15, at 47.

82. Allcott et al., *supra* note 65, at 31, 35.

83. See *id.* at 2-3, 9-10; Allcott & Wozny, *supra* note 71.

84. This is the basic argument of CONLY, *supra* note 28, who emphasizes the need to assess the full set of costs and benefits.

85. See CHRISTOPHER CHABRIS & DANIEL SIMONS, THE INVISIBLE GORILLA: AND OTHER WAYS

count the total number of passes.

After the little movie is shown, the experimenter asks people how many passes they were able to count. Then the experimenter asks: *And did you see the gorilla?* A lot of people laugh at the question. *What gorilla?* Then the movie is replayed. Now that you are not counting passes, you see a gorilla enter the scene, plain as day, and then pound its chest, and then leave. The gorilla (actually a person dressed up in a gorilla suit) is not at all hard to see. In fact, you can't miss it. But when counting passes, many people (typically about half) do miss it.

Behavioral economists have been quite interested in the gorilla experiment, because it shows that people are able to pay attention to only a limited number of things, and that when some of those things are not salient, we ignore them, sometimes to our detriment. Magicians and used-car dealers try to hide gorillas; the same is sometimes true of those who provide credit cards, cell phone service, and mortgages.⁸⁶

Attention is a scarce resource, and attention is triggered by salience; it follows that salience greatly matters. One reason is that System 1 does not closely survey all aspects of social situations, and System 2 may be working hard on other business. When certain features of a product or an activity are not salient, people may disregard them even if they are important, and the result may be individual harm. Complexity and information overload are problems in part because of the importance of salience. When hidden amidst complexity, important features of products and situations might be missed, thus creating real problems. In fact, a lack of salience can be a serious kind of market failure, producing individual and social harm.

Why, for example, do so many people pay bank overdraft fees? One answer is that such fees are not sufficiently salient to people, and some fees are incurred as a result of inattention and neglect. A careful study suggests that limited attention is indeed a source of the problem and that once overdraft fees become salient, they are significantly reduced.⁸⁷ When people take surveys about such fees, they are less likely to incur a fee in the following month, and when they take a number of surveys, the issue becomes sufficiently salient that overdraft fees are reduced for as much as two years.⁸⁸

OUR INTUITIONS DECEIVE US 5-8 (2010).

86. See BAR-GILL, *supra* note 16, at 18-23. Early work by Daniel Kahneman focused on closely related questions. See DANIEL KAHNEMAN, *ATTENTION AND EFFORT* (1973).

87. See Victor Stango & Jonathan Zinman, *Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees* 27-28 (Fed. Reserve Bank of Phila., Working Paper No. 11-17, 2011), http://www.phil.frb.org/phil_mailing_list/research-and-data/publications/working-papers/2011/wp11-17.pdf.

88. *Id.* at 23, 25, 27.

In many areas, the mere act of being surveyed can affect behavior by, for example, increasing use of water-treatment products (thus promoting health) and the acquisition of health insurance; one reason is that being surveyed increases the salience of the action in question.⁸⁹ In the same vein, a field experiment finds that simple textual reminders that loan payments are due have a significant effect on payments – indeed, the same effect as an economic incentive in the form of a twenty-five-percent decrease in interest payments!⁹⁰ A field experiment shows that reminders have a strong effect on people who are due for a dental checkup.⁹¹ Reminders and checklists are effective because they promote salience.

A more general point is that many nontrivial costs (or benefits) are less salient than purchase prices. They are “shrouded attributes” to which some consumers do not pay much attention. Such “add-on” costs may matter a great deal but receive little consideration because they are not salient.⁹² An absence of attention to energy costs, which may be “shrouded” for some consumers, has significant implications for regulatory policy. The clearest such implication involves the importance of providing cost-related information that people can actually understand. In 2011, the Department of Transportation and the EPA produced new fuel-economy labels with this goal in mind; the new labels explicitly draw attention to the economic effects of fuel economy.⁹³

An understanding of the problem of shrouded attributes also helps to identify a potential justification for regulatory standards in the domains of fuel economy and energy efficiency, involving a behavioral market failure. Of course such standards reduce social costs by reducing air pollution and promoting energy security. But from recent rules, the strong majority of the relevant benefits are private; they come from consumer savings.⁹⁴ On standard

89. See Alix Peterson Zwane et al., *Being Surveyed Can Change Later Behavior and Related Parameter Estimates*, 108 PROC. NAT'L ACAD. SCI. 1821, 1825-26 (2011).

90. See Ximena Cadena & Antoinette Schoar, *Remembering To Pay? Reminders vs. Financial Incentives for Loan Payments* 4, 18 (Nat'l Bureau of Econ. Research, Working Paper No. 17,020, 2011), <http://www.nber.org/papers/w17020>. Checklists work for similar reasons. See MARTY MAKARY, UNACCOUNTABLE: WHAT HOSPITALS WON'T TELL YOU AND HOW TRANSPARENCY CAN REVOLUTIONIZE HEALTH CARE 177-78 (2012).

91. See Steffen Altmann & Christian Traxler, *Nudges at the Dentist* (Inst. for the Study of Labor, Discussion Paper No. 6699, 2012), <http://ftp.iza.org/dp6699.pdf>.

92. See Xavier Gabaix & David Laibson, *Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets*, 121 Q.J. ECON. 505, 511 (2006).

93. See Sunstein, *Empirically Informed*, *supra* note 3, at 1373.

94. See NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., DEP'T OF TRANSP., FINAL REGULATORY IMPACT ANALYSIS: CORPORATE AVERAGE FUEL ECONOMY FOR MY 2017-MY 2025 PASSENGER CARS AND LIGHT TRUCKS (2012), http://www.nhtsa.gov/staticfiles/rulemaking/pdf/cafe/FRIA_2017-2025.pdf.

economic grounds, it is not simple to identify a market failure that would justify taking account of such benefits. A plausible argument is behavioral. The basic idea is that such standards might help produce a set of outcomes akin to those that would result if relevant attributes were not shrouded.

This point has not escaped official attention. In explaining the new fuel-economy rules issued in 2012, the Department of Transportation referred to

phenomena observed in the field of behavioral economics, including loss aversion, inadequate consumer attention to long-term savings, or a lack of salience of relevant benefits (such as fuel savings, or time savings associated with refueling) to consumers at the time they make purchasing decisions. Both theoretical and empirical research suggests that many consumers are unwilling to make energy-efficient investments even when those investments appear to pay off in the relatively short-term. This research is in line with related findings that consumers may undervalue benefits or costs that are less salient, or that they will realize only in the future.⁹⁵

So justified, fuel-economy standards are a form of hard paternalism, but they need not question people's ends. The idea is that people want to minimize all relevant costs, and if they are not taking account of some such costs, properly designed fuel-economy standards promote, and do not override, their ends. It is true that if the problem is a lack of attention and salience, the most natural and presumptively appropriate response is disclosure, not a mandate—and on one view, fuel-economy labels, and not a mandate, are the better option. But if such a mandate has benefits far in excess of costs, it would appear to be justified as well.⁹⁶

3. *Unrealistic Optimism*⁹⁷

System 2 is realistic, but System 1 is not.⁹⁸ A great deal of work in

95. *Id.* at 983 (footnote omitted). For a valuable overview, showing the complexity of the underlying issues and the amount that remains to be learned, see Hunt Allcott & Michael Greenstone, *Is There an Energy Efficiency Gap?*, 26 J. ECON. PERSP. 3 (2012). For an important discussion of externalities and internalities, see Allcott et al., *supra* note 65.

96. In the same spirit, see CONLY, *supra* note 28, at 6-12. I do not explore here the question whether fuel-economy standards are the ideal tool or whether other options would be preferable. For relevant discussion, see Allcott et al., *supra* note 65; on the underlying questions, see Allcott & Wozny, *supra* note 71.

97. See Christine Jolls, *Behavioral Economics Analysis of Redistributive Legal Rules*, 51 VAND. L. REV. 1653, 1659 (1998). See generally SHAROT, *supra* note 11 (discussing unrealistic optimism, with particular reference to its neurological foundations).

behavioral psychology and economics suggests that most people are unrealistically optimistic, in the sense that their own predictions about their behavior and their prospects are skewed in the optimistic direction.⁹⁹ Indeed, the tendency toward unrealistic optimism seems to be hardwired.¹⁰⁰ And if people are unduly optimistic about their future behavior, they may select financial packages (say, for credit cards, mortgages, health care plans, and cell phones) that result in significant economic losses.¹⁰¹ In addition, they may run risks (say, by texting while driving) that can lead to serious harm. The most general point is that if people are unduly optimistic, they may fail to take optimal precautions against serious dangers. An obvious response is a disclosure strategy, perhaps including graphic warnings, that helps to counteract unrealistic optimism.¹⁰²

When people imagine their own future, they tend to see it as very good, even if the likely reality is far more mixed.¹⁰³ The “above average” effect is common;¹⁰⁴ many people believe that they are less likely than others to suffer from various misfortunes, including automobile accidents and adverse health outcomes. A study found that while smokers do not underestimate the statistical risks faced by the population of smokers, they nonetheless believe that their personal risk is less than that of the average smoker.¹⁰⁵ Unrealistic optimism is related to confirmation bias, which occurs when people give special weight to information that confirms their antecedent beliefs.¹⁰⁶ To the

-
98. See SHAROT, *supra* note 11; Tali Sharot et al., *How Unrealistic Optimism Is Maintained in the Face of Reality*, 14 NATURE NEUROSCIENCE 1475 (2011) (showing that people do not update on receiving bad information as well as they do when receiving good information).
99. See generally SHAROT, *supra* note 11 (explaining the nature and the sources of skew toward unrealistic optimism); TALI SHAROT, *THE SCIENCE OF OPTIMISM: WHY WE'RE HARD-WIRED FOR HOPE* (2012) (same).
100. Tali Sharot et al., *Selectively Altering Belief Formation in the Human Brain*, 109 PROC. NAT'L ACAD. SCI. 17,058 (2012), <http://www.pnas.org/cgi/doi/10.1073/pnas.1205828109>.
101. See BAR-GILL, *supra* note 16, at 21-26.
102. See Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, 35 J. LEGAL STUD. 199, 215 (2006).
103. See SHAROT, *supra* note 11, at x-xiv.
104. See Neil D. Weinstein, *Unrealistic Optimism About Susceptibility to Health Problems: Conclusions from a Community-Wide Sample*, 10 J. BEHAV. MED. 481, 494-96 (1987). For an interesting complication, showing that people sometimes tend to see themselves as below-average for difficult or unusual tasks, see Don A. Moore & Deborah A. Small, *Error and Bias in Comparative Judgment: On Being Both Better and Worse than We Think We Are*, 92 J. PERSONALITY & SOC. PSYCHOL. 972 (2007).
105. See Paul Slovic, *Do Adolescent Smokers Know the Risks?*, 47 DUKE L.J. 1133, 1136-37 (1998).
106. See David Eil & Justin M. Rao, *The Good News-Bad News Effect: Asymmetric Processing of*

extent that people show this bias, and to the extent that it affects their behavior, they may be led in directions that produce serious welfare losses.

What makes people unrealistically optimistic? How can people maintain such optimism in the face of repeated experiences with reality, which should press them toward greater realism? One reason involves a remarkable asymmetry in how people process information.¹⁰⁷ In brief, people give more weight to good news than to bad news. Tali Sharot and her collaborators find that when people receive information that is better than expected, they are likely to change their beliefs—but when what they learn is worse than expected, their beliefs are more likely to remain constant. In the first stage of the experiment, people were asked to estimate their likelihood of experiencing eighty bad life events (such as robbery and Alzheimer’s disease). In the second stage, they were given accurate information about the average probability for similarly situated people. In the third stage, people were asked to state their view about their personal probability in light of what they had learned.

The central finding is that updating is more likely when people get good news than when they get bad news. More specifically, people were more likely to move their personal probability estimate *upward* when they learned that the population average was above the number they gave than to move their personal probability estimate *downward* when they learned that the population average was below the number they gave. Here, then, is clear evidence of selective updating. The authors conclude that the impact of a learning signal, or new information, “depends on whether [that] new information calls for an update in an optimistic or pessimistic direction.”¹⁰⁸

The authors also studied fMRI data to explore what happens in identifiable regions of the brain—more particularly, the right inferior prefrontal gyrus (IFG), a region of the prefrontal cortex. This is an important question, because the IFG is the region that corrects errors in estimation. Does the IFG react differently to negative and positive information? The answer is yes. The authors’ basic conclusions are technical but worth quoting:

We found that optimism was related to diminished coding of undesirable information about the future in a region of the frontal cortex (right IFG) that has been identified as being sensitive to negative estimation errors. Participants with high scores on trait optimism were worse at tracking undesirable errors in this region than those with low

Objective Information About Yourself, 3 AM. ECON. J.: MICROECONOMICS 114, 116-17 (2011).

107. See Sharot et al., *supra* note 98.

108. *Id.* at 1477. For some compelling evidence of the neural foundations of optimism, and particularly the more ready incorporation of good news than bad news, see Sharot et al., *supra* note 100.

scores. In contrast, tracking of desirable information in regions processing desirable estimation errors (MFC/SFG, left IFG and cerebellum) did not differ between high and low optimists.¹⁰⁹

A subsequent study found that people's ability to incorporate bad news into their judgments can be improved by disrupting the functioning of the left (but not the right) interior frontal gyrus; this disruption eliminates the good news/bad news effect.¹¹⁰ The conclusion, with neural foundations, is that people are unrealistically optimistic, in the sense that they are more responsive to desired than to undesired information—a point that obviously raises challenges for regulatory policy and disclosure requirements in particular. Perhaps the most important point here is that disclosure requirements may turn out to be ineffective with respect to optimistically biased consumers. Any such requirements should be devised so as to reduce that risk; graphic warnings are a possibility here.

4. Problems with Probability

For various reasons, System 1 does not handle probability well. One problem is the availability heuristic. When people use that heuristic, their judgments about probability are affected by whether a recent event comes readily to mind.¹¹¹ If an event is cognitively “available,” people might well overestimate the risk. If an event is not cognitively available, people might well underestimate the risk.¹¹²

In deciding whether it is dangerous to walk in a city at night, to text while driving, or to smoke, people often ask about incidents of which they are aware. While System 2 might be willing to do some calculations, System 1 works quickly, and it is easy and even fairly automatic to use the availability heuristic. Instead of asking hard questions about statistics, it asks easy questions about what comes to mind. “Availability bias” can lead to significant mistakes about the probability of undesirable outcomes.¹¹³ The bias can take the form of either

109. Sharot et al., *supra* note 98, at 1477.

110. See Sharot et al., *supra* note 100.

111. See Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 *COGNITIVE PSYCHOL.* 207, 221 (1973).

112. See Elke U. Weber, *Experience-Based and Description-Based Perceptions of Long-Term Risk: Why Global Warming Does Not Scare Us (Yet)*, 77 *CLIMATIC CHANGE* 103, 107-08 (2006).

113. See Laurette Dubé-Rioux & J. Edward Russo, *An Availability Bias in Professional Judgment*, 1 *J. BEHAV. DECISION MAKING* 223, 233-34 (1988); Paul Slovic, Baruch Fischhoff & Sarah Lichtenstein, *Cognitive Processes and Societal Risk Taking*, in *THE PERCEPTION OF RISK* 32, 37-38 (Paul Slovic ed., 2000).

excessive fear or complacency.

A distinct but related finding is that people sometimes do not make judgments on the basis of the expected value of outcomes, and they may neglect the central issue of probability, particularly when emotions are running high.¹¹⁴ Especially in such cases, people may focus on the outcome and not on the probability that it will occur.¹¹⁵ If there is a small chance of catastrophe—the loss of a child, a fatal cancer—that outcome, rather than the statistical likelihood that it will happen, may dominate people’s thoughts. If there is a small chance of something wonderful—the best vacation ever or a fabulous job opportunity—people’s enthusiasm about that outcome may crowd out the statistics.

Those who sell insurance trade on people’s fear of the worst-case scenarios; so do terrorists, who aim to convince civilians that they “cannot be safe anywhere” in their daily lives. When people are making mistakes about probability, well-designed disclosure strategies, including warnings, could help. Here too, the government would be respecting people’s ends. When officials (or private institutions) correct people’s mistakes about risks, they are affecting means, and helping people to achieve their goals.

II. PATERNALISMS

A. Working Definitions

Do the findings just outlined justify paternalism? The initial task is to produce a working definition of paternalism. Of course paternalism can come from diverse people and institutions. Employers, professors, doctors, lawyers, architects, bankers, rental car companies, and countless others are capable of paternalism. All of these, and many others, may attempt to influence System 1 or to educate System 2, and those efforts, along with social pressures, can greatly affect individual choices. My narrow focus here, however, is on paternalism from government. Though the underlying issues deserve careful attention, and though the discussion here bears on those issues, I do not

114. See George F. Loewenstein et al., *Risk as Feelings*, 127 PSYCHOL. BULL. 267, 280 (2001); Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 YALE L.J. 61 (2002).

115. See Yuval Rottenstreich & Christopher K. Hsee, *Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk*, 12 PSYCHOL. SCI. 185, 185 (2001). For a demonstration that probability is often neglected with respect to things, but not with respect to money (without, however, emphasizing the role of emotions), see A. Peter McGraw, Eldar Shafir & Alexander Todorov, *Valuing Money and Things: Why a \$20 Item Can Be Worth More and Less than \$20*, 56 MGMT. SCI. 816, 827 (2010). For a discussion of emotions and risk, see generally THE FEELING OF RISK: NEW PERSPECTIVES ON RISK PERCEPTION (Paul Slovic ed., 2010).

explore behavioral justifications for paternalism from nongovernmental actors, such as doctors, teachers, lawyers, and employers.¹¹⁶

There are many recent examples of arguable or actual paternalism from public officials. Consider, for example, the controversial decision in 2012, initiated by New York Mayor Michael Bloomberg, to ban the sale (in certain places) of sodas in containers of more than sixteen ounces.¹¹⁷ Mayor Bloomberg sought to reduce obesity, and he believed that the ban would promote that goal. Some people choose drinks in large containers, and Mayor Bloomberg's proposal would not merely influence that choice but make it unavailable. And indeed, much of the negative reaction to the proposal stemmed from the view that it was paternalistic and unacceptable for that reason.¹¹⁸ Why—critics asked and sometimes raged—should Mayor Bloomberg make the decision about the size of soft drink containers, rather than consumers themselves? His proposal was certainly taken as a form of paternalism. (Note that it was a mild form; people could still drink as much as they like; they simply had to buy two containers rather than one. I will return to the question of how best to characterize it below.)

1. Choices and Welfare

It is tempting to suggest that *the government acts paternalistically when it overrides people's choices on the ground that their choices will not promote their own welfare*. But there is an immediate problem with this suggestion. The idea of “overriding” is ambiguous. Government has a series of tools for influencing people. Some of the strongest tools involve incapacitation, with capital punishment and life imprisonment counting as the limiting cases. Insofar as we are speaking of these particular penalties, and of imprisonment more

¹¹⁶. There are important questions, not explored here, about the grounds for distinguishing between paternalism from government and paternalism from nongovernmental actors. One obvious ground involves coercion, but (as discussed in detail below), some forms of government paternalism are not coercive, and some forms of private paternalism can be understood as coercive (for example, when an employer threatens an employee with discharge if he or she does not manage a self-control problem that does not affect others). And while I cannot discuss the complexities here, it is reasonable to wonder whether some of these findings provide new support for paternalistic steps—at least nudges—from those with expertise, including doctors.

¹¹⁷. See, e.g., Kevin Loria, *New York Soda Ban Proposal: Public Hearing Gets Impassioned*, CHRISTIAN SCI. MONITOR, July 24, 2012, <http://www.csmonitor.com/USA/Society/2012/0724/New-York-soda-ban-proposal-Public-hearing-gets-impassioned>.

¹¹⁸. See, e.g., Michael M. Grynbau & Marjorie Connelly, *60% in City Oppose Bloomberg's Soda Ban, Poll Finds*, N.Y. TIMES, Aug. 22, 2012, <http://www.nytimes.com/2012/08/23/nyregion/most-new-yorkers-oppose-bloombergs-soda-ban.html>.

generally, it may be fair to speak of overriding choices. Other tools are more subtle, ranging from monetary penalties, large and small, to the use of education, warnings, default rules, and time, place, and manner restrictions. Even criminal and civil bans are often accompanied by monetary penalties. When the government imposes penalties on certain choices, it puts people who make those choices at some kind of risk or in some kind of jeopardy. Choices are not overridden, strictly speaking. If people are told that they will have to pay a fine if they engage in certain behavior, they remain free to engage in that behavior and to pay the fine. Paternalistic policies may influence rather than override choices.

The unifying theme of paternalistic approaches, however diverse, is that government does not believe that people's choices will promote their welfare, and it is taking steps to influence or alter people's choices for their own good.¹¹⁹ In acting paternalistically, government may be attempting (1) to affect outcomes without affecting people's actions or beliefs; (2) to affect people's actions without influencing their beliefs; (3) to affect people's beliefs in order to influence their actions; or (4) to affect people's preferences, independently of affecting their beliefs, in order to influence their actions. Automatic enrollment would fall in the first category insofar as it affects outcomes; but it need not lead to any change in people's actions.¹²⁰ The power of automatic enrollment stems from the fact that it works on those who are passive.¹²¹ A civil fine would fall in the second category insofar as it affects what people do without affecting their beliefs.¹²² An educational campaign or a set of factual warnings, specifically designed to alter beliefs, would fall in the third. A graphic warning campaign, designed to affect preferences but without necessarily affecting beliefs,¹²³ would fall in the fourth category.

From the standpoint of those who oppose paternalism, all of these effects may be objectionable, but perhaps for different reasons. For example, efforts to affect people's preferences might seem especially insidious except insofar as such efforts are limited to the provision of truthful information. Provision of

119. For a valuable and relevant discussion, bearing particularly on means paternalism, see B. Douglas Bernheim & Antonio Rangel, *Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics*, 124 Q.J. ECON. 51 (2009).

120. Indeed, automatic enrollment often affects outcomes because people take no action at all. See Sunstein, *supra* note 70.

121. Chetty et al., *supra* note 70.

122. It is possible, of course, that a fine could affect beliefs, not just actions. For example, a fine could convey information about the appropriate attitude to have toward an activity or a product, and that information could influence beliefs.

123. To be sure, it might affect people's beliefs (for example, by making people think that the risks are very serious), and influence their behavior for that reason.

such information is certainly a nudge, but it may or may not qualify as paternalistic. I will explore that complex issue below.

2. *The (Important but Troubled) Distinction Between Means and Ends*

I have noted the importance of distinguishing between means paternalism and ends paternalism. In acting paternalistically, government might well accept people's ends but conclude that their choices will not promote those ends. A GPS provides information about how to get from one place to another. People can ignore what the GPS says and try their own route, but if they do so, there is a serious risk that they will undermine their own ends (and people know that). Means paternalists see their proper domain as building on the GPS example. If, for example, people want to make a sensible tradeoff between up-front costs and long-term fuel costs, but sometimes fail to do so (perhaps because long-term costs are not salient), means paternalists might take steps to steer people in the direction of considering all relevant costs at the time of purchase.

We have seen that disclosure is the most natural solution here, but we have also seen that means paternalists would consider a fuel-economy mandate if they could be convinced that such a mandate would promote consumers' ends. The analogy here would be to a GPS that forces cars to take the best or most sensible route—not an entirely attractive idea (what if people enjoy certain scenery, or are nostalgic about longer routes?), but perhaps appealing for some people and at some times and places. The idea of the coercive GPS can be seen as a model and a test for hard paternalism with respect to means.

Ends paternalists have more ambitious goals. They might think, for example, that longevity is what is most important and that even if people disagree, and are willing to run certain risks for reasons they believe to be good and sufficient, paternalists should steer them toward longevity. Or ends paternalists might believe that certain sexual activity is inconsistent with people's well-being, suitably defined, and hence they should not be allowed to engage in that activity. Behavioral economists have not sought to revisit people's ends. They have generally emphasized human errors with respect to means, and hence means paternalism is their principal interest and also my main focus here.

While the distinction between means paternalism and ends paternalism captures something important, it raises a number of questions, and the line between the two is not always sharp. Some of the most straightforward cases of means paternalism involve shrouded attributes, optimism bias, and availability bias. Suppose that people want a refrigerator that will perform well and cost as little as possible. If government ensures that people have accurate information about cost, it is not revisiting their ends in any way. Indeed, it is not even

acting paternalistically, in the sense that it is informing people's choices, rather than (independently) influencing them. The same can be said if people underestimate the risks of distracted driving or of smoking. If the government corrects people's unrealistic optimism, or counteracts the effects of the availability heuristic to produce an accurate judgment about probability, it is respecting their ends, and we might not want to characterize its action as paternalistic at all.

So too if, for example, people are ignoring certain product attributes because those attributes are shrouded. If those attributes would matter to people if they attended to them, then efforts to promote disclosure do not question people's ends. Of course there may be hard questions here in determining whether people are in fact ignoring shrouded attributes (as opposed to not caring about them), but thus far, at least, there is no problem of ends paternalism, and indeed there might not be paternalism at all. If the relevant steps are harder—if they involve economic incentives designed to discourage the relevant behavior, or flat bans—then they would qualify as paternalistic. But they would seem to count as means paternalism if they are designed only to ensure that people achieve their own ends.¹²⁴

Even in the apparently easy cases, however, there are complications. Consider a fuel-economy label, designed to inform people of the cost over a year or a five-year period of particular cars. If the government provides this information through a vivid letter grade—say, an “A” or a “B,” as was in fact proposed¹²⁵—it is not merely providing people with facts. To be sure, this is not the most aggressive form of paternalism about either means or ends. But formal grades might be taken as a form of paternalism not merely about means, but also about ends, insofar as government is singling out the particular variable of fuel economy and attempting to focus people's attention on that variable, as opposed to numerous other variables that would remain ungraded. And indeed, the government declined to require letter grades in part on the ground that such grades might be taken, wrongly, to suggest that the government was giving “all things considered” grades to cars.¹²⁶ But I am making a different point here: even if this risk did not exist, a fuel-economy grade could be taken to be paternalistic, and to involve a degree of paternalism about ends as well as means, insofar as it would focus and heighten people's attention with respect to one of innumerable features of cars. The government does not, after all, give serious consideration to requiring letter grades with respect to speed, or acceleration, or brightness of color, or stylishness, or

124. See CONLY, *supra* note 28, at 149–80.

125. See Sunstein, *Empirically Informed*, *supra* note 3, at 1373.

126. See *id.*

coolness (actual or perceived).

Even without letter grades, any fuel-economy label itself has at least a degree of paternalism, certainly about means and indeed about ends as well, insofar as it isolates fuel economy, rather than other imaginable features of cars, for compulsory display. Consider a thought experiment or perhaps a little science fiction. We should be able to agree that the government would focus only on means, and would not be paternalistic, if it could have direct access to all of people's internal concerns and provide them with accurate information about everything that concerns them. And perhaps in the fullness of time, government, or the private sector, will be able to do something like that. But insofar as the government is being selective, it is at least modestly affecting people's ends, and perhaps intentionally so.

Of course people want to save money; that is one of their ends. But the government chose a fuel-economy label, rather than an acceleration label or a coolness label, for a reason—to focus consumer attention on that particular feature of cars. (To be sure, it is also possible that fuel economy is more shrouded than other features, but that is hardly self-evident.) To these points we might add the more familiar one, which is that any disclosure requirement has to be framed in a certain way, and the choice of frame may well affect people's decisions and even their ends.

It is reasonable to say that the government would be focused solely on means if it provided people with accurate information about everything that they cared about. In that event, disclosure would not be paternalistic at all. It would be means focused, and it would not attempt to influence choices except insofar as it would promote accurate beliefs, which is not a paternalistic endeavor. But if the government frames a disclosure policy with the purpose and effect not only of informing but also of influencing people's choices, it is engaging in a form of soft paternalism—not only about means, but also about ends, insofar as it is attempting to affect them. And if the government's disclosure policy is selective, in the sense that it requires disclosure with respect to one attribute (that people care about) but not others (that people also care about), it is again engaging in a form of soft paternalism about means and also ends, insofar as it is attempting to affect them—unless it can be shown that the selected attribute is, distinctly, one on which people now lack and need information.

But we should not be too fussy or clever here, and we really should avoid tying ourselves into conceptual knots. If framing or selectivity is at work, there may be a form of ends paternalism, but it is likely to be of a very modest kind. If the characteristic is one that people antecedently do care about—like money—then it is fair to say that any paternalism is at least centrally about means, and that the intrusion on people's ends is modest and possibly even

incidental.

In the domain of procrastination and time inconsistency, however, the distinction between means paternalism and ends paternalism is more troubled still. In addressing those problems, are paternalists addressing means or ends? If ends, at what time? At Time 1, the person sought to smoke, to drink, and to eat a lot; at Time 2, the (same) person wishes that none of these choices had been made.¹²⁷ To know whether a paternalistic intervention is about means or about ends, we may have to identify the level of generality at which people's ends are to be described. If the end is "for life to go well," then all forms of paternalism, including the most ambitious, seem to qualify as means paternalism, since they are styled as means to that most general of ends. But if the end is very specific—"To buy this product today!" or "To smoke this cigarette right now!"—then many and perhaps all forms of paternalism qualify as ends paternalism. If ends are described at a level of great specificity, there may be no such thing as means paternalism.

In the hard cases of procrastination and time inconsistency, the best solution may be to decline to answer the "means or ends" question directly, on the ground that it is not tractable, and instead to ask about people's aggregate welfare over time, on the theory that aggregate welfare (taking all relevant values into account) is the end that people really do care about. If an effort to overcome unjustified procrastination promotes people's welfare on balance, it responds to a behavioral market failure and hence is plausibly justified, at least on welfare grounds. The word "plausibly" is important; there are many objections, and I will get to them in due course. And of course public officials may face formidable problems in deciding what promotes aggregate welfare over time¹²⁸—a point that argues in favor of soft rather than hard paternalism, and one to which I will return.

3. *The (Important but Troubled) Distinction Between Hard and Soft*

Let us dispense with the idea of "overriding" choices and emphasize the different tools that paternalistic officials are using. We can imagine actions of government that attempt to improve people's own welfare by threatening to *imprison* those who make certain choices. We can also imagine actions of government that attempt to improve people's own welfare by threatening to *fine* those who make certain choices. If the government imposes criminal or

127. I am bracketing here any questions about personal identity over time. See generally DEREK PARFIT, *REASONS AND PERSONS* (1986).

128. See CONLY, *supra* note 28, at 1-16, for the plausible suggestion that some cases are really not so difficult by this measure.

civil fines on those who smoke marijuana, refuse to buckle their seatbelts, or gamble, and if it does so because it disagrees with people about what would promote their own welfare, it is acting paternalistically.

There is of course a continuum here between paternalistic actions that impose high costs and paternalistic actions that impose low costs. A small monetary fine—of, say, five cents—falls within the definition of paternalism, but it may not have a significant effect on behavior. Note, however, that some sanctions have *expressive functions* and may be effective for that reason, even if the actual size of the sanction is small.¹²⁹ A modest criminal fine (say, for smoking, failing to buckle one’s seatbelt, texting while driving, or gambling) may have a large deterrent effect. A paternalistic intervention with such a sanction, however modest, might be found highly objectionable by those who abhor paternalism. Note in addition that even very small costs—say, a five-cent charge for a bag at a grocery store—may have a significant effect on behavior.¹³⁰ And indeed, a careful analysis shows such an effect, in part because of the power of loss aversion.¹³¹ If such small costs are imposed in order to protect people against their own bad or harmful choices, they count as paternalistic.

In fact, it might be best to understand paternalistic interventions in terms of a continuum from hardest to softest, with the points marked in accordance with the magnitude of the costs (of whatever kind) imposed on choosers by choice architects. On this view, there is no sharp or categorical distinction between hard paternalism and soft paternalism; all we have are points along a continuum. But we should agree that there is a significant difference between, say, a severe criminal ban on smoking marijuana and a nominal civil fine, and between a prison sentence for failing to buckle your seatbelt and a graphic educational campaign offering vivid warnings.

Under this approach, a statement that paternalism is “hard” would mean that choice architects are imposing large costs on choosers, whereas a statement

129. See Robert A. Kagan & Jerome H. Skolnick, *Banning Smoking: Compliance Without Enforcement*, in *SMOKING POLICY: LAW, POLITICS, AND CULTURE* 69, 72 (Robert L. Rabin & Stephen D. Sugarman eds., 1993) (finding high compliance in part because of the expressive function of law).

130. See Josh Kizler, *Update: The Efficacy of Washington, D.C.’s Bag Fee*, [PLASTICBAGLAWS.ORG](http://plasticbaglaws.org/update-the-efficacy-of-washington-d-c-s-bag-fee) (Mar. 23, 2011), <http://plasticbaglaws.org/update-the-efficacy-of-washington-d-c-s-bag-fee>; see also DAN ARIELY, *PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS* 49-63 (rev. ed. 2009) (discussing the concept of “free” and the significant effect it has on people’s decisionmaking).

131. See Tatiana A. Homonoff, *Can Small Incentives Have Large Effects? The Impact of Taxes Versus Bonuses on Disposable Bag Use* (Mar. 27, 2013) (unpublished manuscript), http://www.princeton.edu/~homonoff/THomonoff_JobMarketPaper.

that paternalism is “soft” would mean that the costs are small. And under this approach, all costs, material or nonmaterial, would count, and to assess the degree of hardness, we would inquire into their magnitude. For example, psychic costs, as produced by graphic warnings, could move an intervention along the continuum toward hard paternalism, as long as those costs turned out to be high. Nudges would count as soft paternalism because and insofar as they impose no or very small costs on choosers.

There are significant advantages in seeing a continuum here rather than a categorical distinction. But if a categorical distinction is what is sought, we should focus on the existence of material costs. On this approach, we would understand the term “hard paternalism” to refer to *actions of government that attempt to improve people’s own welfare by imposing material costs on their choices*. By contrast, the term “soft paternalism” would refer to *actions of government that attempt to improve people’s own welfare by influencing their choices without imposing material costs on those choices*.

If the government engages in an advertising campaign designed to convince people to exercise more than they now do, it is engaging in a kind of soft paternalism. If the government requires employers automatically to enroll workers in health insurance plans, or requires warnings to accompany certain products, soft paternalism is involved. Soft paternalism is libertarian insofar as it does not impose material costs on people’s choices. (Of course, material costs are being imposed in all of these cases; the focus is on whether those costs are being imposed on the choices of end-users.) We can understand soft paternalism, thus defined, as including nudges, and I will use the terms interchangeably here.

In a careful and highly illuminating book, Riccardo Rebonato offers a provocative and different definition of libertarian paternalism, or nudges:

Libertarian paternalism is the set of interventions aimed at overcoming the unavoidable cognitive biases and decisional inadequacies of an individual by exploiting them in such a way as to influence her decisions (in an easily reversible manner) towards choices that she herself would make if she had at her disposal unlimited time and information, and the analytic abilities of a rational decision-maker (more precisely, of *Homo Economicus*).¹³²

This definition is useful, but it is imprecise in three respects. First, the universe of nudges is far broader than the definition suggests. Soft paternalism includes interventions (such as warnings and default rules) that may be

132. REBONATO, *supra* note 28, at 6.

helpful, but that need not specifically counteract biases and decisional inadequacies. Second, the word “counteracting” is better than “exploiting.” Nudges can counteract biases (such as unrealistic optimism) without exploiting anything. Third, the words “easily reversible” are imprecise, because they could capture (for example) small civil penalties, even though they do not count as libertarian.

Emphasizing the idea of a continuum, however, we should recognize that approaches that impose (high) psychic costs, and thus target System 1, may have a greater effect, and in that sense turn out to be less soft, than approaches that impose (low) material costs. Moreover, an approach that does not impose high material costs may have a major effect on choices. Indeed, it may greatly affect both beliefs and actions, and hence make all the difference. People may well change their behavior when psychic costs are high even if material costs are close to zero. An emphasis on material costs may be useful for purposes of taxonomy, but it should not be taken to suggest that such costs are all that matter, or even that tools that impose such costs are the most influential ones in the toolbox.

4. A Very Quick Summary

Summarizing these various points, we can imagine the following possibilities, with illustrative examples:

	MEANS PATERNALISM	ENDS PATERNALISM
SOFT PATERNALISM	Fuel-economy labels	Automatic enrollment in particular political party
HARD PATERNALISM	Fuel-economy standards	Criminal ban on same-sex relations

Where behavioral market failures justify corrective action, the government should be inclined to stay in the upper-left quadrant, unless strong empirical justifications, involving relevant costs and benefits, support a more aggressive approach. Recall the first law of behaviorally informed regulation, which is that in the face of behavioral market failures, nudges are generally the right response. Moreover, those who emphasize behavioral market failures would seek to avoid both quadrants on the right-hand side.

5. On Welfare

My account of paternalism raises an immediate question. What counts as

people's welfare? Does it mean happiness, narrowly conceived? Might it include whatever makes lives good and meaningful, even if happiness, strictly speaking, is not involved? I will return to these questions. For now, and to keep the focus on the issue of paternalism, I am going to understand the term "welfare" very broadly (and in a way that clearly separates the capacious idea of welfare from the narrower one of utility). Let us also notice the importance of distinguishing between "welfare" from the standpoint of the chooser and "welfare" from the standpoint of the paternalist.

With respect to the chooser, let us understand the term to refer to whatever choosers think would make their lives go well.¹³³ Choosers might, for example, care about the taste, amount, and nutritional content of food and drink. They might be happy to eat a lot of high-calorie foods, every day, simply because they enjoy them so much. (They might dislike or even hate calorie labels, on the ground that they detract from the enjoyment.) Or they might care not only about the economic benefits of fuel-efficient cars but also about the environment.

Their principal concerns might be religious; they might believe that fidelity to God's will is what is necessary to make their lives go well. When they think about their own lives, they may want to make choices that benefit other people—not only their friends and families, but strangers as well. They may want their lives to be meaningful, not merely full of pleasure, and they might sacrifice material and other benefits to achieve that goal. They strike their own balance; different people will choose differently. They may or may not enjoy exercising or smoking. They may or may not care a lot about health effects or aesthetics.

With respect to the paternalist, we can understand "welfare" in the same way, to refer capaciously to whatever the paternalist thinks would make choosers' lives go well. The paternalist might believe that choosers have the right ends, but that some kind of action is needed to ensure that they actually achieve those ends (perhaps because of the operation of System 1). Alternatively, the paternalist might believe that choosers have the wrong ends—perhaps choosers do not focus enough on health, or sexual abstinence, or on what makes life meaningful, or on obedience to God's will—and that some kind of response is needed, with respect to actions or beliefs, to ensure that the right ends are achieved. Though paternalists might have any number of views about what would make people's lives go well, my focus throughout is on paternalists who respect people's own views about their ends, and who seek to ensure that their decisions promote those ends.¹³⁴

133. See CONLY, *supra* note 28, for an instructive discussion.

134. See the treatment of perfectionism in *id.* at 100-25.

My working definition of paternalism does not include government efforts to prevent people from harming others—as, for example, in the case of assault or theft, or air pollution. There is nothing paternalistic about preventing people from beating you up, stealing your car, or making the air unsafe to breathe. Nor does the definition include government efforts to produce certain familiar and widely held social goals; consider laws designed to protect endangered species, or to prevent discrimination on the basis of race, sex, disability, and sexual orientation. None of these is fundamentally rooted in paternalistic considerations.¹³⁵ By contrast, the definition includes government efforts to override people’s judgments about whether it is best for them to drink alcohol, to gamble, to drive while talking on their cell phones, or to eat a dozen chocolate peanut butter cookies before, during, or after dinner.

True, and importantly, some of these cases may involve harm to others. If you drive while talking on your cell phone, you might endanger other people, and perhaps a restriction could be defended for that reason. If you are making yourself drunk or even sick, you might affect others. In the cases just described, it is possible that regulation can be justified on grounds that have nothing to do with paternalism. To see some of the complexities here, recall recent rules that require increases in fuel economy. Such rules produce substantial social benefits by reducing air pollution and by increasing energy security; producing these benefits does not involve paternalism. But as we have seen, the strong majority of the benefits of such rules come from private fuel savings,¹³⁶ and producing these benefits might well be thought to involve paternalism. To get clear on the underlying issues, let us put third-party effects entirely to one side.

If we begin with this definition, the central concern about paternalistic interventions, elaborated most famously in John Stuart Mill’s *On Liberty*,¹³⁷ is that people must remain free to choose as they see fit. The focus is on preventing certain action by the state, unless harm to others is involved. We should be able to see that while the principal objection is to ends paternalism, means paternalism can raise serious problems as well. Even in the face of

135. It is possible, however, that paternalistic justifications could be offered for these laws, involving the effects of discrimination on choosers. There is also a possibility of auto-paternalism, as for example when people seek to bind themselves, perhaps through law. See JON ELSTER, *ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY* (1984). With the exception of a few brief remarks, I do not explore auto-paternalism here, though it does have an obvious relationship to some responses to behavioral market failures from which people might be attempting to protect themselves.

136. See NAT’L HIGHWAY TRAFFIC SAFETY ADMIN., *supra* note 94.

137. MILL, *supra* note 1. Mill famously focused not merely on paternalistic government but on a wide set of influences on individual behavior, including influences that come from the private sphere. My own focus here is of course narrower.

behavioral market failures, why should public officials be authorized to interfere with people's judgments about the best means to promote their ends? Mightn't they err as well, and possibly more damagingly? These are important questions, but one of my principal goals here is to suggest that insofar as the Millian view neglects the existence of behavioral market failures, and the wide range of behavioral findings about human errors, it points in exactly the wrong direction.¹³⁸

B. The Paternalist's Large Toolbox

To know whether and what kind of paternalism is involved, and to get clearer on the underlying concepts, we need to be more specific about the set of tools that government might use. Consider some possibilities:

1. Government says that no one may smoke cigarettes and that the sanction for smoking cigarettes is a criminal penalty—of \$500.
2. Government says that no one may smoke cigarettes and that the sanction for smoking cigarettes is a criminal penalty—of \$0.01.
3. Government says that no one may smoke cigarettes and that the sanction for smoking cigarettes is a civil fine—of \$500.
4. Government says that no one may smoke cigarettes and that the sanction for smoking cigarettes is a civil fine—of \$0.01.
5. Government does not say that no one may smoke cigarettes, but instead imposes a tax on cigarette purchases—a tax of \$2.00 per pack.
6. Government does not say that no one may smoke cigarettes, but instead creates a program that provides a financial subsidy to smokers who quit for six months—a subsidy of \$500.
7. Government does not say that no one may smoke cigarettes, but instead engages in a vivid, frightening advertising campaign, emphasizing the dangers of smoking.¹³⁹
8. Government does not say that no one may smoke cigarettes, but instead requires packages to contain vivid, frightening images,

¹³⁸ In the same general vein, see CONLY, *supra* note 28, at 1-15. Note, however, that Conly argues in favor of coercive (or "hard") forms of paternalism, which I treat very cautiously here. We do agree that the master concept involves an assessment of costs and benefits (including the frustration felt by those whose choices are influenced by paternalists).

¹³⁹ See *Tips from Former Smokers*, CTRS. FOR DISEASE CONTROL & PREVENTION, <http://www.cdc.gov/tobacco/campaign/tips/resources/videos> (last updated Jan. 3, 2013).

- emphasizing the dangers of smoking.¹⁴⁰
9. Government does not say that no one may smoke cigarettes, but instead engages in a public education campaign designed to make smoking seem deviant, antisocial, or uncool.
 10. Government does not say that no one may smoke cigarettes, but instead engages in a truthful, fact-filled educational campaign disclosing the dangers of smoking.
 11. Government does not say that no one may smoke cigarettes, but instead requires packages to provide truthful information disclosing the dangers of smoking.
 12. Government does not say that no one may smoke cigarettes, but instead requires cigarette sellers to place cigarettes in an inconspicuous place, so that people will not happen across them and must affirmatively ask for them.
 13. Government does not say that no one may smoke cigarettes, but instead requires cigarettes to be sold in small containers, each having no more than five cigarettes. (Cigarette packs usually have twenty cigarettes now.)

Those who begin with the definition I have offered should acknowledge that all of these cases are not the same. If we are focused on leaving freedom of choice unaffected by government, and use the definition offered above, approaches (1) through (5), involving penalties, would count as forms of paternalism. Approach (6) might be seen as more difficult. Is it paternalistic to subsidize behavior? Does paternalism include not merely penalties but also subsidies? What about selective subsidies, as in, for example, a decision to allow recipients to use food stamps to pay for almost all food and drink, but not soda or chocolate bars? (In 2011, Mayor Bloomberg asked the United States Department of Agriculture for permission not to allow food stamps to be used to pay for soda; the Department denied the petition.¹⁴¹) Insofar as a subsidy is designed to influence a person's choices on the ground that those choices would not promote his or her welfare, it should be counted as paternalistic.

By contrast, disclosure of truthful information is not ordinarily understood as paternalistic. As we have seen, the basic reason is that disclosure

¹⁴⁰. See Required Warnings for Cigarette Packages and Advertisements, 76 Fed. Reg. 36,628 (June 22, 2011) (to be codified at 21 C.F.R. pt. 1141). The rule was invalidated in *R.J. Reynolds Tobacco Co. v. FDA*, 696 F.3d 1205 (D.C. Cir. 2012), but upheld in *Discount Tobacco City & Lottery, Inc. v. United States*, 674 F.3d 509 (6th Cir. 2012).

¹⁴¹. Michael Howard Saul, *Mayor's Soda Plan Fails*, WALL ST. J., Aug. 20, 2011, <http://online.wsj.com/article/SB10001424053111903596904576518902332775160.html>.

requirements are meant to inform, not to displace, people's understanding of which choices will promote their welfare. But we have also seen several complexities here. First, disclosure of information will often affect that understanding, especially if—and because—it is selective. Second, the framing of information much matters,¹⁴² and any disclosure requirement will inevitably include a certain kind of framing. It may be disputed whether a given disclosure requirement is simply informing choices; some forms of disclosure can certainly fall within the category of soft paternalism.

What about approaches (7) and (8), involving the use of vivid, frightening images? I have emphasized that psychic costs, no less than material costs, can alter behavior. Some people might think that efforts to frighten people, and thus to go beyond mere disclosure of facts and to grab the attention of System 1, can be taken as a form of (soft) paternalism. Under the definition I have offered, it is more than plausible to hold this view. Indeed, at least one court has drawn a distinction of this kind for First Amendment purposes, suggesting that compelled disclosure of facts is different from, and more acceptable than, compelled graphic warnings.¹⁴³ Any efforts to stigmatize a product, and to do so through emotional appeals, might be seen as imposing a psychic or affective cost on purchase or use. Imposition of affective costs is paralleled by the creation of affective benefits, which could come, for example, by efforts to portray certain activities, such as exercise or eating vegetables, in a positive light; such approaches could also be characterized as soft paternalism.

Approaches (12) and (13) also involve forms of soft paternalism. If officials put a product in an inconspicuous place, and if their goal is to discourage its purchase, they are steering people in a certain direction because they distrust people's own judgments about what would promote their welfare. No monetary penalty is involved, but time and effort must be expended to find the relevant goods. And if government requires a product to be sold in small containers so that people will consume less of it, it is behaving paternalistically insofar as it is making it harder for them to make the choices that they prefer. True, many people may prefer that private or public institutions impose such costs, and some or many smokers may themselves share that preference because they would like to quit—but the point remains.

¹⁴². For example, an effect might be framed as a gain or as a loss. For an excellent collection, see *PERSPECTIVES ON FRAMING* (Gideon Keren ed., 2011).

¹⁴³. See *R.J. Reynolds Tobacco Co. v. FDA*, 845 F. Supp. 2d 266, 272 (D.D.C. 2012), *aff'd*, 696 F.3d 1205.

III. AGAINST PATERNALISM: WELFARE

Let us put the definitional issues to one side. Why might paternalism be objectionable? There are two time-honored reasons; both have firm roots in Mill's own argument. The first involves welfare. The second involves autonomy.

Because we are focusing here on paternalism from the government, we should observe that there are special concerns in that context. On a familiar view, it may not be unacceptable if a private employer acts paternalistically in an effort to protect its employees, or if a credit card company does so in an effort to protect its customers. Because of the operation of the free market, harmful, insulting, or unjustified paternalism will ultimately be punished.¹⁴⁴ But if government acts paternalistically—to improve health, to lengthen lives, to save money—some people believe that the question is different. Certainly it is true that many of the most prominent objections to paternalist interventions (decrying “elitism,” “government overreach,” or “the nanny state”) have a great deal to do with the distinctive social role of particular paternalists: those who work for the government.

To be sure, there are political constraints on the actions of a paternalistic government, at least in a democratic society, and on an optimistic view, those constraints will sharply limit the occasions for harmful or unjustified paternalism. But that view may be too optimistic, especially in light of the fact that some forms of paternalism are not salient or highly visible,¹⁴⁵ and the associated fact that well-organized private groups, with their own interests at stake, may wish to move public policy in their preferred directions.

Whatever the origins of the objections to paternalistic government, the force of those objections should depend on whether paternalism, from government, threatens to reduce people's welfare (broadly understood) or to intrude on people's autonomy. It is not unreasonable to fear that the risks of a paternalistic government are more serious than the risks of private paternalism. But if those risks come to fruition, the underlying concerns involve welfare, autonomy, or both. Many of the concerns about paternalistic government focus on the idea of “legitimacy,” but in this context, at least, it is possible that the term is a placeholder, or perhaps even a mystification, rather than a freestanding concept.¹⁴⁶

^{144.} This may be too optimistic in some cases. See BAR-GILL, *supra* note 16, at 26–32 (exploring the extent to which market pressures may encourage exploitation of behavioral biases).

^{145.} See *infra* Section V.A.

^{146.} It may be a freestanding concept insofar as government action is challenged as insufficiently democratic, but I am stipulating that there is no such objection here. If not, an objection

I focus on welfarist concerns in this Part and turn to autonomy in the following Part. I begin with a set of apparently powerful welfarist objections to paternalism and then turn to responses, focusing on the fact that as an empirical matter, paternalism can in fact increase people's welfare. I suggest that a great deal depends on context and that the objections cannot be shown to be convincing in the abstract. In fact, the problem is that these objections operate at too high a level of generality, potentially making them into a form of chest-thumping. The best that can be said is that there is an intelligible rule-consequentialist argument against paternalism, both hard and soft, but the argument is unhelpful insofar as choice architecture (and hence a form of paternalism) is inevitable. Insofar as paternalism can be avoided, the rule-consequentialist objection depends on empirical claims that are unlikely to be true, at least in important contexts.

A. Five Welfarist Objections: An Antipaternalist's Quintet

1. *Information.* Suppose that we care about people's welfare, understood broadly to capture how well their lives are going. Suppose that we believe that people's lives should go as well as possible. We might insist that individuals know best about what will make their lives go well and that public officials are likely to err. Such officials might be mistaken about what people's ends are, and they might also be mistaken about the best means of achieving those ends. People might really enjoy running, sleeping, having sex, singing, jumping, smoking, drinking, gambling, or (over)eating. They might have their own views about how, exactly, to go about enjoying those activities.

So long as they are not harming others, people should be allowed to act on the basis of their own judgments, because those judgments are the best guide to what will make their lives go well. A central argument, applicable to any kind of paternalism (soft or hard, means or ends), is that errors are more likely to come from officials than from individuals. Public officials lack the information that individuals have.

2. *Competition.* In a free economy, companies compete with one another, and people are free to choose among a range of options. If a refrigerator is not cold enough and if it costs a lot of money to operate, it will not do well in the market. Companies will produce better refrigerators that cost less. If cars have poor fuel economy and end up costing a lot over time, companies will compete to increase fuel economy. If important features of products are shrouded, and

from "legitimacy" would seem to add nothing. The question is whether the action at issue makes people's lives better (understanding that idea very capaciously) or intrudes on their autonomy.

are bad, they will eventually be revealed. Some consumers may be fooled or tricked, but in the long run, the process of competition will help a great deal.

Here is a major problem: paternalistic approaches may freeze the process of competition. Especially if they are hard rather than soft, they may impair the operation of a competitive process that produces a mixture of diverse products that are well suited to diverse tastes and circumstances. We can thus identify a Hayekian challenge to paternalism.¹⁴⁷ Even if public officials are armed with knowledge of behavioral market failures, and even if they are public-spirited, they will do far worse than free markets, which can produce a wide range of products and rapid responses to changing tastes and needs.

I have emphasized that participants in the market are able to counteract all of the problems identified here. Companies themselves can help promote self-control. They can reveal shrouded attributes. They can counteract unrealistic optimism. They can promote an accurate understanding of probability. As technologies evolve, such correctives should be increasingly available. For every behavioral market failure, it may soon be possible to say, with the old Apple commercial, “There’s an app for that.”

3. *Learning*. It is true that people err, and their errors can impair their welfare. But mistakes are often productive. Life is a movie, not a snapshot, and people can learn from what goes wrong. We should not freeze people’s frames. On one view, government ought not to short-circuit the valuable process of learning-by-doing. That process greatly increases human welfare. Indeed, people become better choosers as a result. If people make mistakes about diets, drinks, love, or investments, they can obtain valuable lessons, and those lessons can make their lives go much better.

Perhaps there is no reasonable concern about efforts to ensure that people’s choices are well informed—at least if those efforts do not discourage people from learning on their own. But if people are defaulted into a certain savings or health care plan, rather than asked to choose such a plan on their own, learning is less likely to occur. In a sense, soft paternalism can infantilize its citizens by preventing such learning, and reduce liberty in the process.

For those who emphasize the value of learning, it might seem best to call for *active choosing* rather than default rules.¹⁴⁸ In many areas, government might

¹⁴⁷. See F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 524-26 (1945).

¹⁴⁸. See, e.g., Gabriel D. Carroll et al., *Optimal Defaults and Active Decisions*, 124 Q.J. ECON. 1639 (2009). For a vigorous suggestion that soft paternalism will impede learning, see Wright & Ginsburg, *supra* note 28; on the underlying issues, see Bruce Ian Carlin, Simon Gervais & Gustavo Manso, *Libertarian Paternalism, Information Production, and Financial Decision-Making* (June 6, 2012) (unpublished manuscript), <http://faculty.fuqua.duke.edu/~sgervais/Research/Papers/LibertarianPaternalism.WP.pdf>.

dispense with default rules and instead require people to make choices on their own. Perhaps this approach can also be counted as a form of soft paternalism insofar as it is steering people toward active decisions, when people have not actively decided in favor of active decisions! But even if so, it will be congenial to those who emphasize both learning and choice.

4. *Heterogeneity*. Human populations are highly diverse in terms of tastes and values, and one size is unlikely to fit all. In many contexts, an effort to impose a single size will reduce welfare on balance. With respect to diet, savings, exercise, credit cards, mortgages, cell phones, health care, computers, and much more, different people have divergent tastes and situations, and they balance the relevant values in different ways.¹⁴⁹ The same is true with respect to tradeoffs between the present and the future. People have diverse ends, and they choose diverse means. Young people will make different tradeoffs from old people. It is hardly irrational to value the present more than the future, and different discount rates can reasonably be chosen by people who are in different life circumstances. There may be no self-control problem if people decide to enjoy today and tomorrow, even if the consequence is not ideal for the day after.

To be sure, we should not rely on abstractions here. We have to investigate the details, and relevant empirical questions, to know whether and how heterogeneity matters. If people are required to buckle their seatbelts or to wear motorcycle helmets, and if they are forbidden to text while driving, it is at least imaginable that no matter how diverse the population, the welfare of the overwhelming majority of people will be increased as a result. On plausible assumptions, they will live longer and safer lives, and they will not lose a lot.

The admittedly serious problems with one-size-fits-all approaches should not be taken to suggest that one size never fits all. (With respect to one-size-fits-all approaches, universal skepticism is itself a one-size-fits-all approach, and a bad one.) But the simple fact of human diversity suggests that if government prescribes a certain outcome, and departs from people's own sense of what is best, human welfare might be reduced rather than increased. For those who find these points convincing, soft paternalism has significant advantages, but insofar as it steers all people in the same direction, it raises problems of its own.

We should emphasize, however, that paternalists, both hard and soft, might be able to manage the problem of heterogeneity by avoiding one-size-fits-all approaches and by attempting more personalized approaches.

149. See MILL, *supra* note 1, at 121-22, for a version of this argument. Related discussion can be found in Glaeser, *supra* note 28, at 135-42, which emphasizes the ability of those in the private sector to balance relevant values and to incorporate new information.

Personalized paternalism is likely to become increasingly feasible over time.¹⁵⁰ We can, for example, imagine highly personalized default rules, attempting to specify diverse default rules for people in different circumstances. Such approaches might draw on available information about people's own past choices or about which approach best suits different groups of people, and potentially each person, in the population. Personalized default rules might be based on demographics; a default savings plan for someone who is thirty would be different from a default savings plan for someone who is sixty. Alternatively, personalized default rules could be very narrowly targeted. If enough information is available about someone's past choices or personal situation, we could design, for that person, default rules with respect to health insurance, privacy, rental car agreements, computer settings, and everything else.¹⁵¹

Personalized default rules would reduce the problems posed by one-size-fits-all approaches, and, in principle at least, personalized approaches might even eliminate those problems. To be sure, the design of personalized paternalism raises serious technical challenges, and it remains unclear whether it could fully respect heterogeneity, especially in light of the fact that people's preferences and situations change over time.¹⁵² In many ways, personalization does appear to be the wave of the future, but there is still a serious argument for active choosing, which would eliminate paternalism.¹⁵³

5. *Public choice (including behavioral public choice)*. It should not be necessary to emphasize that public officials have their own biases and their own motivations. With respect to efforts to defend paternalism, this point raises two separate problems. The first involves public choice theory: official judgments about welfare may be influenced by the interests of powerful private groups.¹⁵⁴ No one can deny that, at some times and places, official judgments have been distorted because of the power of such groups. The second problem is that even if they are well motivated, officials are human too, and there is no reason to think that they are immune from the kinds of biases that affect ordinary people.¹⁵⁵

We can go further. At the moment, there is no such field as behavioral

150. See Sunstein, *supra* note 70, at 27-32.

151. *Id.*

152. For detailed discussion, see *id.* at 27-32.

153. *Id.*

154. See, e.g., GORDON TULLOCK, ARTHUR SELDON & GORDON L. BRADY, *GOVERNMENT FAILURE: A PRIMER IN PUBLIC CHOICE* (2002).

155. See Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 *STAN. L. REV.* 683 (1999).

public choice theory, but there should be. Behavioral public choice theory would supplement the standard accounts by exploring the extent to which public officials go wrong because they make the very errors outlined here.¹⁵⁶ True, we might expect System 2 to have a great deal of authority in government, simply because government has in its employ many people whose business it is to calculate the consequences of alternative consequences of action, and thus to affirm the primacy of System 2.¹⁵⁷ As we have seen, a large virtue of technocrats in government—specialists in science, economics, and law—is that they can help overcome some of the errors that might otherwise influence public as well as private judgments.¹⁵⁸

Nonetheless, System 1 is not irrelevant in the public domain. There is no question that availability bias can play a role in public arenas.¹⁵⁹ Recent unfortunate events might lead people to think that a problem is more serious than it actually is, and the absence of such events might lead people to neglect real problems because no recent misfortune comes to mind. Officials are hardly immune from availability bias: if a bad outcome has occurred in the recent past, it becomes highly salient, and it may affect ultimate decisions. Indeed, officials may be subject to a kind of *anticipatory availability bias*: their anticipation of a terrible outcome, and of being blamed for such an outcome, can affect their judgments. Self-interested private groups may aggravate the problem by repeatedly drawing official attention to bad outcomes (or may promote complacency by repeatedly drawing official attention to the absence of bad outcomes).

Behavioral public choice theory would explore these problems in great detail. On the basis of the discussion thus far, we should be able to identify its ingredients. For every bias identified for individuals, there is an accompanying bias in the public sphere. This point offers serious cautionary notes about paternalism, whether it addresses people's means or ends.

If these points are put together, the central problem with paternalism is that it will, in the end, make people's lives go worse. Because it allows for greater flexibility, soft paternalism is less objectionable than hard paternalism, but all of the foregoing points might be brought to bear against paternalism of any kind.

156. See *id.* for one set of examples.

157. For a general discussion, see Cass R. Sunstein, *The Office of Information and Regulatory Affairs: Myths and Realities*, 126 HARV. L. REV. (forthcoming 2013).

158. See Cass R. Sunstein, *Cognition and Cost-Benefit Analysis*, 29 J. LEGAL STUD. 1059 (2000); Sunstein, *supra* note 157.

159. See Kuran & Sunstein, *supra* note 155.

B. Welfare: Normative Issues

1. *Welfare revisited.* We should begin by acknowledging that to some people, the arguments just sketched will fall on deaf ears. They will seem puzzling, question-begging, even perverse.

If we focus on welfare, we have to return to the initial question, which is what it means for lives to go well. Mill was a utilitarian, and he focused on increasing people's utility. If that is our focus, and if we understand utility in a certain way, we might well be drawn to his antipaternalist conclusion (subject to serious empirical challenges, which should by now be evident and which are taken up in more detail below). But if we have a different understanding, and if we believe that in order to go well, lives must take a particular form, the objection to paternalism will seem badly confused and perhaps unintelligible. Indeed, ends paternalism, no less than means paternalism, will seem legitimate, and not even slightly undermined by the antipaternalist arguments made thus far.

Suppose, for example, that we are not focused on utility and that we start with a theological view that emphasizes obedience to God's commands and that does not put a high premium on freedom of choice. If so, what some people deplore as paternalism will seem to others the natural and appropriate way to ensure that people's lives go well. In an illuminating book, Jonathan Haidt emphasizes the existence of plural and diverse foundations for moral commitments.¹⁶⁰ The antipaternalist view depends on accepting some such commitments and rejecting others.

To be more specific: for those who begin with an emphasis on *purity*, the arguments from welfare and autonomy will have little force. Suppose we think that for a life to be pure, and therefore good, people must refrain from certain activities, including gambling, smoking, drinking, and overeating, and that other activities, such as sex, can occur only subject to certain restrictions. If a life goes well if and only if it is pure (in a relevant sense), then hard paternalism will seem to be the right course, and the antipaternalist argument will face an obvious (and devastating) problem.

For that argument to get off the ground, we have to start with what some will find contentious views, to the effect that human lives can go well in many different ways, and that people are generally the best judges of how to make their own lives go well. That (broadly Millian) view will in turn help fuel the belief that individuals are usually the best judges of what it means for their lives to go well.¹⁶¹ In my view, there is a great deal to be said for that belief

^{160.} See Haidt, *supra* note 61.

^{161.} See the discussion of perfectionism in Conly, *supra* note 28, at 100-25. I do not explore here

(subject to empirical reservations), but it must be acknowledged that many others are doubtful.

2. *Choices everywhere (and mostly made for us)*. There is an independent point. The welfarist objections neglect the extent to which countless decisions are already made for us by both public and private institutions. Most of us do not decide how to make a car safe, or where to put stop signs, or how and whether to test foods to reduce the risk of disease, or whether and what antibiotics should be allowed on the market, or how to build airplanes and railroads. If we had to make all of the choices that affect us, we would be immediately overwhelmed, and our welfare would be decreased as a result. People make only a very small fraction of the decisions that actually affect them. In important cases, and for good reasons, we can opt out in various ways, but we are on a specific track if we do nothing at all.

True, we can participate in free markets, and we can also vote. For these reasons, we do have some degree of control over many of the underlying choices, at least if “we” are taken in the aggregate. But for each of us, the degree of control is modest. I will return to this point in more detail in the context of the discussion of autonomy.

C. *Welfare: Empirical Problems*

Let us put the deepest issues to one side and simply notice that even if we are concerned about welfare, and even if we are inclined to think that individuals are generally the best judges of how to make their own lives go well, the word “generally” is important. With that qualification, we can see that the objection to paternalism depends on some empirical judgments. Those judgments might be wrong—not (on the behaviorally informed view I am exploring) because it is important or desirable to revise or revisit people’s ends, but because people may select the wrong means to promote their own ends.¹⁶²

Do people’s choices in fact promote their welfare? The answer is knowable, at least in principle, and it is being tested, with mixed results. We learn more every day. The findings discussed in Part I suggest that behavioral market

the strand in Mill that emphasizes the importance of “experiments in living.” See Elizabeth S. Anderson, *John Stuart Mill and Experiments in Living*, 102 ETHICS 4 (1991). Appropriate responses to behavioral market failures, of the kind outlined here, would not seem to run afoul of Mill’s emphasis on such experiments. To be sure, some actions that appear to demonstrate self-control problems might be seen as experiments in living, but disclosure requirements and warnings allow such experiments to continue. It is hard to see how fuel-economy requirements, or energy efficiency mandates, plausibly jeopardize experiments in living, even though these are hard forms of paternalism.

¹⁶². I acknowledge and bracket the difficulties in this distinction. See *supra* Subsection II.A.2.

failures are far from uncommon, and as we have seen, they supplement the standard (welfarist) justifications for government action. If, for example, people pay too little attention to the long term, and enjoy short-term benefits at the expense of significant long-term costs, then a concern for welfare might require, rather than forbid, certain forms of paternalism (potentially including hard forms). If people procrastinate, and if System 1 is the reason, then their failure to alter the status quo may be a mistake, with possibly bad and even dangerous consequences.

To avoid misunderstanding: the point here is emphatically not that System 2 should be in charge and that what appeals to System 1 does not much matter.¹⁶³ There is no claim that life must be dry, chocolate-free, and long. The point is instead that people make mistakes about what they would enjoy—or to return to our terms, that System 1 makes mistakes about what it will find appealing.

A growing literature explores the difference between “decision utility” and “experienced utility”—the difference between the utility that we think we will get when we make a decision and the utility we actually experience after that decision has been made.¹⁶⁴ The central finding is that at the time of decision, people think that they will obtain a certain amount of utility, or welfare, from certain products or activities—but they sometimes err. We might think that a very expensive car would be a joy to own, but we might get used to that car, and, after a while, we might not get a lot of pleasure from it. The most serious cases are those in which we make choices that greatly endanger our health, and shorten our lives, but the phenomenon is much broader than that.

True, people learn, and true, pleasure is hardly the only thing that people do or should care about. We choose certain activities not because they are fun or joyful, but because they are right to choose, perhaps because they are meaningful. People want their lives to have purpose; they do not want their lives to be simply happy.¹⁶⁵ People sensibly, even virtuously, choose things that they will not in any simple sense “like.”¹⁶⁶ For example, they may want to help

163. See REBONATO, *supra* note 28.

164. See, e.g., KAHNEMAN, *supra* note 6; Dunn et al., *supra* note 12, at 115. For some important cautionary notes about the ability of even close friends to know what people will like, see JOEL WALDFOGEL, SCROOGENOMICS: WHY YOU SHOULDN'T BUY PRESENTS FOR THE HOLIDAYS (2009).

165. Daniel J. Benjamin et al., *What Do You Think Would Make You Happier? What Do You Think You Would Choose?*, 102 AM. ECON. REV. 2083, 2085-86 (2012).

166. See Niklas Karlsson, George Loewenstein & Jane McCafferty, *The Economics of Meaning*, 30 NORDIC J. POL. ECON. 61, 62 (2004); Peter A. Ubel & George Loewenstein, *Pain and Suffering Awards: They Shouldn't Be (Just) About Pain and Suffering*, 37 J. LEGAL STUD. S195, S207 (2008).

others even when it is not a lot of fun to do that. A survey suggests that people's projected choices are *generally* based on what they believe would promote their subjective well-being—but that sometimes people are willing to make choices that would sacrifice their happiness in favor of promoting an assortment of other goals, including (1) promoting the happiness of their family, (2) increasing their control over their lives, (3) increasing their social status, or (4) improving their sense of purpose in life.¹⁶⁷

But I am speaking of cases in which people are really focusing on what they will like, and what they experience as a result of their choices is not what they hope and expect. Their choice does not make them happy. It might even make them seriously ill (or dead). In this sense, they select the wrong means to their own ends. The unlovely technical term here is “affective forecasting errors.”¹⁶⁸ In fact, we can easily imagine cases in which people choose certain actions explicitly on the ground that those actions are likely to be meaningful—and they turn out to be wrong. Affective forecasting errors are paralleled by many other kinds of errors, including those that involve other goods that people care about.

People might, for example, believe that a certain decision would increase the happiness of family members, but they might be entirely wrong in that belief. Consider the finding that people often choose bad Christmas presents for those they love, thus producing billions of dollars in deadweight losses every year.¹⁶⁹ People might believe that a certain outcome will increase their status or their sense of purpose or meaning, but they might be wrong in that belief as well. I am not aware of any empirical work on “meaningfulness forecasting errors,” but I forecast that there will be some.

Here is a simple but striking example of the possibility that hard paternalism can actually increase people's welfare. We would ordinarily expect people to be worse off if public officials make it more expensive for them to purchase goods that they want. If government tells you that you have to spend more to buy a computer, a book, a lamp, or a pair of shoes, your life will not be better. But there may be exceptions. More specifically, *cigarette taxes appear to make smokers happier*.¹⁷⁰ To the extent that this is so, it is because smoking makes smokers less happy. When smokers are taxed, they smoke less and may even quit—and they are better off as a result. (Historical note: In 2009, President Obama, himself a former smoker, signed a law that increased the tax

167. Benjamin et al., *supra* note 165, at 2085.

168. See Gilbert et al., *supra* note 12, at 618.

169. See WALDFOGEL, *supra* note 164.

170. See Gruber & Mullainathan, *supra* note 73, at 2.

on cigarettes by \$0.62, from \$0.39 to \$1.01.¹⁷¹)

This finding is most puzzling if we are inclined to think that people's decisions always increase their welfare. If we believed that, we would hardly expect to find that people are better off if their choices are taxed and in that sense discouraged. For various reasons, including its addictive nature, smoking is of course a highly unusual activity, and it is generally true that if we want to make people better off, the best approach is hardly to increase the price of goods that they want. But smoking may not be unique. It is not unimaginable that people would be happier as a result of other taxes on goods that they choose. Consider taxes on foods that cause obesity.¹⁷² Whether or not such taxes can be justified on balance (and this is a complex question on which I am expressing no view here), there is an intelligible argument for them. In fact, we might be able to see some taxes as analogous to commitment strategies, meant to address externalities, in which people agree to put obstacles in their own way; recall the tale of Ulysses and the Sirens.¹⁷³

The broader point is that in some cases, there is real space between anticipated welfare and actual experience. The space suggests that if welfare is our guide, the antipaternalist position will run into serious problems, especially in cases that involve serious risks to life or health. True, it might be rescued if we have good reason to think that whatever the errors made by individuals, they are less frequent, and less damaging, than the errors made by public officials.¹⁷⁴ But that question also requires empirical investigation, and, in the abstract, the answer is not clear.

D. Imaginable Worlds and Rule-Consequentialist Antipaternalism

It should now be obvious that the welfarist arguments against paternalism, whatever their form, depend on empirical assumptions and perhaps even hunches. We could certainly imagine a world—call it Millville—in which the best approach, from the standpoint of welfare, is to let people decide as they see fit, and to impose a flat ban on government efforts to influence their

171. Wendy Koch, *Biggest U.S. Tax Hike on Tobacco Takes Effect*, USA TODAY, Apr. 3, 2009, http://www.usatoday.com/money/perfi/taxes/2009-03-31-cigarettetax_N.htm.

172. See, e.g., Jeff Strnad, *Conceptualizing the "Fat Tax": The Role of Food Taxes in Developed Economies*, 78 S. CAL. L. REV. 1221 (2005); Lucia Reisch & Wencke Gwozdz, *Smart Defaults and Soft Nudges: How Insights from Behavioral Economics Can Inform Effective Nutrition Policy* (May 31, 2012) (unpublished manuscript) (on file with author) (discussing the uses and limits of nudges and outlining and apparently approving of steps in Denmark, Hungary, and France to tax certain foods).

173. See ELSTER, *supra* note 135.

174. See Glaeser, *supra* note 28; Wright & Ginsburg, *supra* note 28.

decisions.¹⁷⁵ We could also imagine a world—call it Benthamville—in which any such flat ban would be far too crude, if our goal is to increase welfare, and in which we would want to make some distinctions designed to maximize welfare by, for example, authorizing paternalism when the risks of widespread private error are especially high (and the risks of government error low), and forbidding paternalism when those risks are low (and the risks of government error high). We could even imagine a world—call it Nirvana—in which public officials could be trusted, so that the space for paternalism would be significantly increased. We could easily imagine a world in which the form of paternalism much mattered, so that criminal penalties would be strongly disfavored, but subsidies and taxes would be acceptable, or in which criminal penalties, subsidies, and taxes would be strongly disfavored, but in which nudges would be entirely acceptable.

Which of these worlds is our own? Reasonable people differ. Some of the strongest objections offer a kind of explicit or implicit rule-consequentialism.¹⁷⁶ Those who make such objections acknowledge that people err and that it is possible, in principle, that public officials could promote people's welfare. But they suggest that if we want to promote welfare on balance, we should not go case by case, but should instead adopt a rule or at least a presumption against paternalism, whether hard or soft, and whether focused on means or ends.

The rule-consequentialist position would be supported with the following questions: Aren't public officials human as well? Who will monitor them? Who will nudge them? What about the value of private learning? We might reiterate that public officials are hardly invulnerable to the cognitive errors described here. Even if their distinctive role makes their System 2 likely to be unusually engaged, their susceptibility to private pressure may raise distinctive concerns.

To be sure, we should not use the public choice problem as a kind of all-purpose battering ram or trump card, and it is possible to identify cases in which people are better off if government is authorized to act paternalistically. But according to one view, the risks outweigh the potential gains. On that view, we should adopt a general rule against paternalism on rule-consequentialist grounds, not because the general rule always leads in good directions, but because it is far safer, and far better, than a case-by-case approach.

175. I am bracketing here the inevitability of choice architecture and the fact that a number of choices are made for people already.

176. See Glaeser, *supra* note 28, which has a rule-consequentialist flavor, but which is qualified through a recognition that (optional) nudging is justified in identifiable cases.

E. Choice Architecture and Inevitable Nudges

We have seen an immediate objection to the rule-consequentialist suggestion, and it cannot be repeated often enough, simply because it is so often ignored (and so please forgive the italics): *choice architecture is inevitable*. The social environment influences choices, and it is not possible to dispense with a social environment.¹⁷⁷ Default rules are omnipresent, and they matter. Do we have an opt-in design or an opt-out design? Whenever there is an answer, there is an effect on outcomes.

Does this mean that paternalism is also unavoidable? Suppose that we use the definition set out above, so that paternalism is involved when public officials do not believe that people's choices will promote their welfare, and hence are taking steps to influence or alter people's choices for their own good. If so, we might think that while choice architecture cannot be avoided, it is possible to avoid paternalism. Perhaps choice architects—at least if they are working for the government—can self-consciously reject or avoid any efforts to influence or alter choices. Government officials might respect people's choices, and the choice architecture that is established by the private sector, and attempt to avoid any independent effects of their own. It is true that officials can work to minimize such effects. But some choice architecture is likely to be in place from government, and no such architecture is entirely neutral.¹⁷⁸ If officials are setting up websites or cafeterias, or producing forms or applications of various kinds, it is likely that their decisions will have some effect on what people select.

The rule-consequentialist objection would therefore have to be more refined. It would be that government should avoid paternalism whenever it is feasible to do so. And it must be agreed that warnings and educational campaigns can be abandoned, or stopped before they start. It is also true that officials can work, in many cases, to eliminate default rules and to rely on active choosing. If we think that public officials are overwhelmingly likely to err, or to be poorly motivated, we might think that the risks of official action outweigh the benefits. This thought cannot be rejected in the abstract. In some times and

177. This point holds whether the social environment is a product of self-conscious designers or some kind of invisible-hand mechanism. There can be choice architecture without choice architects. For a valuable discussion, see Edna Ullmann-Margalit, *Invisible-Hand Explanations*, 39 *SYNTHESE* 263 (1978).

178. For a discussion, see THALER & SUNSTEIN, *supra* note 5, at 1-4; and Richard H. Thaler, Cass R. Sunstein & John P. Balz, *Choice Architecture*, in *BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY*, *supra* note 10, at 428, 428-31. I am bracketing here the potential effects of the kinds of choice architecture that are established by the basic rules of contract law, property law, tort law, and criminal law.

places, it is undoubtedly correct.

A welfarist approach would not rely on that abstract possibility. It would ask about costs and benefits. At least in the United States, what we know about nudges and their effects¹⁷⁹ makes it extremely hard to defend the rule-consequentialist objection, even in its more refined form. Would we really be better off if government did not inform people of the risks of smoking and of driving without seatbelts? Of texting while driving? Of the nutritional content of food? Should government blind itself to what it knows about behavioral market failures?

A more general point involves the relationship between System 1 and System 2. Many of the errors and biases discussed here are driven by System 1. Public officials do not exactly lack a System 1—far from it—but as I have noted, much of their job is to rely on System 2, by assessing costs and benefits and by devoting careful thought to options and consequences. We need not be naïve about this process to agree that at least in well-functioning democracies, the power of System 2, in the public domain, operates as a valuable safeguard.

Certainly there are complications here, and we might have to make some distinctions among political actors. Within the executive branch, the emphasis on analysis of costs and benefits can operate as a System 2 safeguard against mistakes.¹⁸⁰ Recall the analogy to speaking in a foreign language; cost-benefit analysis can itself be seen as such a language. At the same time, elected officials, including those in Congress, may or may not be relying on careful analysis. Often they do so, of course. But in at least some cases, their own intuitive reactions, and those of their constituents, may drive judgments about policy and perhaps even legislation.

It is true that the structure of the national legislature was designed to promote careful deliberation. James Madison wrote that the Senate was “to consist in its proceedings with more coolness, with more system and with more wisdom, than the popular branch.”¹⁸¹ The same idea is reflected in a much-quoted exchange between Thomas Jefferson and George Washington. When Jefferson asked why the Constitutional Convention had created a Senate, Washington noted that “we pour legislation into the senatorial saucer to cool it.”¹⁸² Nonetheless, such cooling does not always occur.

179. See Sunstein, *Empirically Informed*, *supra* note 3.

180. See Exec. Order No. 13,563, 76 Fed. Reg. 3821 (Jan. 21, 2011) (directing agencies to catalogue costs and benefits and to ensure, to the extent permitted by law, that the benefits justify the costs); Sunstein, *supra* note 157.

181. JAMES MADISON, NOTES OF DEBATES IN THE FEDERAL CONVENTION OF 1787, at 83 (Adrienne Koch ed., 1984) (statement of James Madison).

182. For one account, see *Senate Legislative Process*, U.S. SENATE, <http://www.senate.gov>

My basic conclusion is that while there are strong welfarist objections to paternalism, those objections have no force when some kind of paternalism is inevitable, and in any case they depend on controversial normative judgments and on empirical conjectures that are sometimes right and sometimes wrong. If the goal is to promote welfare, we will have to pay careful attention to context. Taken singly or in combination, the objections should not be taken as trump cards. When there is a behavioral market failure, and when it is causing serious harm, it is implausible—a form of evidence-free dogmatism—to say that a public response is off-limits, especially but not only if it takes the form of soft paternalism. If welfare is our guide, it is necessary to take behavioral market failures seriously.

IV. AGAINST PATERNALISM: AUTONOMY

Suppose that we believe that freedom of choice has a special and independent status. Liberty, and not welfare, might be our guide. We might insist that people have a right to choose and that government cannot legitimately intrude on that right even if and when it does in fact know best. If people want to buy twenty-four-ounce soda bottles, energy-inefficient refrigerators, or cars that have poor fuel economy, they are entitled to do just that. If they want to gamble or smoke, to spend their money rather than save it, or to exercise just once a year (perhaps the day after New Year's?), the government has no business intervening, even if those choices cause them harm.

On this view, people should not be regarded as children; they should be treated with respect.¹⁸³ They should be seen as ends, not means. If government substitutes its own judgments for those of choosers, it violates these principles. The real problem, on this view, is that all forms of paternalism, including those that grow out of an understanding of behavioral market failures, endanger liberty. Here too, however, we should make a distinction.

A. Autonomy: The Thin Version

The thin version of this position suggests that freedom of choice is an ingredient of welfare, and when we decide what government should do, we need to take account of the harmful effect, on welfare, of interfering with that

[/legislative/common/briefing/Senate_legislative_process.htm](#) (last visited Dec. 5, 2012).

183. See REBONATO, *supra* note 28; Wright & Ginsburg, *supra* note 28. Conly discusses and rejects the objection from respect. CONLY, *supra* note 28, at 33-42.

freedom.¹⁸⁴ On this view, people often dislike having their choices overridden, punished, or even significantly influenced; they experience a loss in welfare, and possibly a serious one. People want to choose for themselves. When the government tells people that they have to save money, or cannot text while driving, or have to buckle their seatbelts, it may be making them less happy, and possibly frustrated and angry. The welfare loss that comes from eliminating choices may be large, and it has to be taken into account.

Note in particular that the thin version raises questions about default rules, not merely about mandates and bans. Under a default rule, people are automatically placed in a certain situation unless they opt out. If people like to choose, perhaps it would be better to have a regime of active choosing,¹⁸⁵ avoiding default rules altogether and simply asking people what they want. One advantage of this approach is that people's choices may well promote their welfare better than those of public officials. Another advantage is that choosing promotes learning. Yet another advantage – and the one that I am emphasizing here – is that many people like choosing as such.

In some contexts, the thin version is certainly correct. When people enjoy freedom of choice, and suffer when it is overridden, that loss must be counted in the overall assessment. If people want to select their own retirement plans, or hate the idea of being forced to buckle their seatbelts or to wear motorcycle helmets, public officials must consider those desires.

It is important, however, to see that on the thin version, freedom of choice is relevant but perhaps not decisive. The welfare gain of the paternalist action may outweigh the welfare loss. (True, the measurement issues are formidable here.) Perhaps people would feel frustrated, but perhaps their lives would be much longer and much better. It is also important to see that in some contexts, people do not enjoy freedom of choice and would much prefer not to have to spend time on the question at all.¹⁸⁶ Especially in complex and unfamiliar domains, active choosing can be a burden, not a benefit. There are also issues about the extent to which active choosing increases or decreases satisfaction with ultimate choices. I will return to these points.

B. Autonomy: The Thick Version

The thick version of this position stresses not that freedom of choice is part of welfare, but that it is an end in itself and thus decisive, or at least a very

184. See REBONATO, *supra* note 28, at 138.

185. See Carroll et al., *supra* note 148.

186. See CONLY, *supra* note 28, at 90-94; SHEENA IYENGAR, *THE ART OF CHOOSING* (2010).

weighty matter, to be overridden only for the most compelling reasons.¹⁸⁷ To treat people with respect,¹⁸⁸ and as ends rather than mere means, government cannot override that form of freedom even if doing so would, in fact, make people happier or better off in a relevant sense. On the thick version, imposing costs on those who exercise freedom of choice, or steering people in government's preferred directions, is presumptively unacceptable as such.

Many of the most deeply felt objections to paternalism, strong or weak, are based on an intuition or judgment of this kind. Those objections often take the form of a question: By what right can government legitimately attempt to alter the choices of free adults? (Is this question asked by System 1? Welfarists think so, or more precisely, they believe that System 2 is asking in its capacity as System 1's lawyer or public relations manager; they insist that this question is a rhetorical flourish and therefore unhelpful. I will return to this issue as well.)

C. *Thin, Again*

Begin with the thin version. Suppose that freedom of choice is part of what people care about. Suppose, too, that if people are denied freedom of choice, they will suffer a loss in welfare, in part because they feel frustrated and mistreated. To the extent that this is so, there will be a legitimate point, on grounds of welfare, against hard paternalism (and perhaps soft as well).

1. *Balancing*. As I have noted, it is important to see that the point may not be decisive.¹⁸⁹ Perhaps people are only mildly distressed to lose freedom of choice; perhaps they consider such freedom a burden, at least in new and unfamiliar contexts. If, for example, the question is the precise content of a retirement plan, some people might be glad if the employer selects a plan that meets their needs (subject of course to opt out), and the issue does not seem much different if the employer is the federal or state government. Or perhaps the welfare gain from influencing or even overriding choice is very large, because people would choose in a way that would cause them serious harm.

187. Those who emphasize autonomy tend to allow override when compelling reason exists. For an overview, see Larry Alexander, *Deontology at the Threshold*, 37 SAN DIEGO L. REV. 893, 898-901 (2000). For an emphasis on the importance of autonomy, see Wright & Ginsburg, *supra* note 28.

188. For discussion and criticism, see CONLY, *supra* note 28, at 9: "It is not disrespectful to accurately estimate someone's abilities, and to respond to those appropriately. If anything, coercive paternalism manifests respect for the value of human lives by trying to help people live fruitful lives in which they are able to achieve their own ultimate goals."

189. See *id.* at 10 ("[A]doption of paternalism will require that we undertake a cost-benefit analysis as to whether or not it is worth interfering in people's behavior, and one major element of cost is the feelings of those who are imposed upon.").

(Consider bans on suicide.) The thin version of the autonomy argument raises a highly relevant question, one that may argue against hard paternalism and in favor of active choosing. But it is an empirical issue whether that question, properly answered, raises a serious problem for a proposed act of paternalism.

2. *The social background.* Actually, there is a deeper problem, to which I briefly referred in the context of welfare: all of us could, in principle, make far more decisions than we do in fact. Every hour of every day, choices are implicitly made for us, by both private and public institutions, and we are better off and more autonomous as a result.

Most of us do not have to make choices about what a refrigerator or an alarm clock should look like, or how best to clean tap water, or how to fly an airplane, or what safety equipment should be on trains, or what medicine to take if we have strep throat, or whether certain antibiotics should be available, or where highways and street signs are located. Time is limited, and some issues are complex, boring, or both. If we did not benefit from an explicit or implicit delegation of choice-making authority, we would be far worse off, and in an important sense less autonomous, because we would have less time to chart our own course. Autonomy depends on a social background, whose basic ingredients we need to be able to take for granted. Without that background, and if active choosing were required for everything, our autonomy would quickly evaporate.

Esther Duflo, one of the world's leading experts on poverty, says the following:

[W]e tend to be patronizing about the poor in a very specific sense, which is that we tend to think, "Why don't they take more responsibility for their lives?" And what we are forgetting is that the richer you are the less responsibility you need to take for your own life because everything is taken care for you. And the poorer you are the more you have to be responsible for everything about your life [S]top berating people for not being responsible and start to think of ways instead of providing the poor with the luxury that we all have, which is that a lot of decisions are taken for us. If we do nothing, we are on the right track. For most of the poor, if they do nothing, they are on the wrong track.¹⁹⁰

190. Susan Parker, *Esther Duflo Explains Why She Believes Randomized Controlled Trials Are So Vital*, CENTER FOR EFFECTIVE PHILANTHROPY BLOG (June 23, 2011), <http://www.effectivephilanthropy.org/blog/2011/06/esther-duflo-explains-why-she-believes-randomized-controlled-trials-are-so-vital>. Duflo develops these ideas in detail in her 2012 Tanner Lectures. See Duflo, *supra* note 41. A relevant and valuable discussion is Shah et al.,

Dufló's central claim is that people who are well off do not have to be responsible for a wide range of things, because others are making the relevant decisions, and to their benefit. We need not focus in particular on the disparity between rich and poor to see that as a matter of fact, decisions are taken for all of us by both private and public institutions. Of course it is exceedingly important that we can revisit (many of) those decisions if we do not like them. But if we had to make all relevant choices in the first instance, we would be worse off—and far less free—as a result.

D. Thick, Again

The strong version of the autonomy argument does not turn on empirical questions, and it is, in a sense, a showstopper. If people have to be treated as ends rather than as mere means, and if this principle requires government not to influence private choices, there is not a lot of room for further discussion. We might be forced to acknowledge that if we accept a certain view of autonomy, actions that fall in the category of hard paternalism—such as those that impose significant criminal or civil penalties on behavior that does not harm anyone else—are presumptively out of bounds. In such cases, those who believe in autonomy will insist that government needs an extremely strong reason to interfere with private choices. But does government really treat people as mere means when it does not impose material costs and merely attempts to promote better choices? Is there anything insulting or demeaning about automatic enrollment in savings and health care plans, subject to opt-out? Which nudges, and which forms of libertarian paternalism interfere with autonomy, rightly understood?¹⁹¹

Perhaps we can agree that in some cases, the interest in autonomy does justify a preference for active choosing rather than a default rule. But in all contexts? What if the area is highly technical, and people would consider active choosing a burden rather than a benefit, and a default rule would reduce the number and magnitude of errors?

Consider a stronger and perhaps reckless response to those who invoke autonomy.¹⁹² On one view, what really does and should matter is welfare, for

supra note 40.

191. I explore below some possible answers involving the behavioral difficulties of supposedly easy reversibility.

192. A view of this sort is defended and elaborated by Joshua Greene. See, e.g., Joshua D. Greene et al., *Cognitive Load Selectively Interferes with Utilitarian Moral Judgment*, 107 *COGNITION* 1144 (2008) (“[O]ur theory associates utilitarian moral judgment (approving of harmful actions that maximize good consequences) with controlled cognitive processes and associates non-utilitarian moral judgment with automatic emotional responses. Consistent with this

which autonomy claims are best understood as a heuristic. More precisely, autonomy is what matters to System 1, but on reflection, the real concern, vindicated by System 2, is welfare. On this view, objections from autonomy are far from pointless, and for one reason: when we vindicate autonomy, we generally promote welfare. But on this view, it is much better, and much less crude, to focus directly on welfare.

These points raise the possibility that we need a kind of behavioral economics for judgments of morality, and not merely judgments of fact.¹⁹³ Consider an analogy: the availability heuristic helps to produce assessments of probability, and it generally works well. When we learn of an incident in which certain actions produced serious harm, we update our probability judgments, and the updating is sensible. Use of the availability heuristic can be seen as a kind of rough-and-ready Bayesianism. The problem is that use of the availability heuristic can also go badly wrong, leading to exaggerated fears. The same problems arise for many moral precepts, which generally work well but can lead us in bad directions.

Consider the heuristic: do not lie. The prohibition on lying is a heuristic, and for most of us, System 1 has thoroughly internalized it, so that whatever the circumstances, a lie produces distress and perhaps even a physical reaction (such as a rapid heartbeat and sweaty hands). To the extent that lie detectors tend to work, that is why. But System 2 knows that lying is sometimes acceptable and even obligatory, as when it is necessary to save a life. (“No one is with me,” says the parent of the kidnapping victim to the kidnapper, as she brings a police officer to the scene.) What I am suggesting, very tentatively, is the possibility that the objection from autonomy may be a heuristic and that what we really should care about is welfare. When we respect autonomy, we generally promote welfare, and when we think about paternalism, perhaps welfare is what matters.

I have acknowledged that this suggestion may be reckless, and there are many possible responses to it. On a competing view, System 1 speaks in terms of welfare, and System 2 is able to make the case for autonomy.¹⁹⁴ Certainly many pages have been devoted to the elaboration of what autonomy

theory, we find that a cognitive load manipulation selectively interferes with utilitarian judgment. This interference effect provides direct evidence for the influence of controlled cognitive processes in moral judgment, and utilitarian moral judgment more specifically.”); Joshua D. Greene, *The Cognitive Neuroscience of Moral Judgment*, in *THE COGNITIVE NEUROSCIENCES* 987 (Michael S. Gazzaniga ed., 4th ed. 2009).

193. See Joshua D. Greene, *Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains*, 11 *TRENDS COGNITIVE SCI.* 322 (2007); Cass R. Sunstein, *Moral Heuristics*, 28 *BEHAV. & BRAIN SCI.* 531 (2005); sources cited *supra* note 192.

194. I am noting this possibility, but I am not aware of any evidence to support it.

requires.¹⁹⁵ Those who invoke autonomy do not by any means rest content with their intuitions; they justify their views. Perhaps human beings have immediate intuitions about both welfare and autonomy, and the real question is how best to evaluate those intuitions.¹⁹⁶ As we learn more about the operation of the brain, it may be possible to make progress on these issues, and there is some suggestive (if preliminary) evidence that System 1 is distinctly associated with an emphasis on the importance of autonomy.¹⁹⁷

To say the least, those who emphasize autonomy are not likely to be convinced by what I am calling my reckless suggestion. Let us take the objection on its own terms. It bears repeating that in ordinary life, choice architecture ensures that we do not have to make countless imaginable decisions. Of course, we are allowed to participate in markets and to vote, and in these ways, we are able to influence choice architecture of multiple kinds. In many contexts, we can opt out. But our autonomy is promoted, not undermined, by the existence of choice architecture ensuring that we will be just fine if we do not make particular decisions, so that we are freed up to concentrate on those matters that most concern us. If we had to make far more decisions, our autonomy would be badly compromised, because we would be unable to focus. This is not a point about heuristics. It is a point about the limitations of time, interest, and concern.¹⁹⁸

If we emphasize autonomy, we are not likely to object to efforts to ensure that people are adequately informed. Active choosing may be best, but we will not object strongly to the use of default rules, certainly not if such rules reflect the likely choices of informed people.¹⁹⁹ So long as freedom of choice is maintained, and government does not impose significant costs on those who seek to go their own way, autonomy is not undermined.

Suppose that we stipulate that governments should not treat people merely as means. Nudges and information disclosure do not run afoul of that prescription.²⁰⁰ In particular, disclosures facilitate autonomous decisionmaking

195. See, e.g., F.M. KAMM, *INTRICATE ETHICS: RIGHTS, RESPONSIBILITIES, AND PERMISSIBLE HARM* 78-90 (2007).

196. For relevant discussion, see 1 DEREK PARFIT, *ON WHAT MATTERS* (2011).

197. See *supra* note 192.

198. See CONLY, *supra* note 28, at 90 (“I hate the time that self-regulation takes from things that I actually am interested in While some people write as if every time a freedom were taken from us we kick and scream and feel deprived, others, more realistic, recognize that the responsibility for making such choices is a burden, and one that we are often quite willing to give up.”); Duflo, *supra* note 41.

199. For discussion, see Sunstein, *supra* note 70.

200. The difficulties of apparently easy reversibility, discussed *infra* Section V.B, do raise a complication.

by allowing individuals to make fully informed decisions about their own ends.

The chief response to those who invoke autonomy, then, is that reasonable responses to behavioral market failures ought not to raise serious concerns, certainly not if they take the form of nudges. And in some cases, even harder forms of paternalism may not run afoul of autonomy concerns, at least if they respect people's ends. Consider fuel-economy standards that are based on fleet-wide averages, and that allow a diverse array of vehicles to continue to be available to consumers.

E. An Accounting

I have explored a number of tools and approaches here. Most of them do not raise serious problems from the standpoint of those who offer the strongest objections to paternalism.

1. *Disclosure: just the facts.* Disclosure policies, mandated or authorized by law, should be designed sensibly, so as to inform people rather than to be unduly complex or unintelligible. Such efforts promote freedom of choice. The new Food Plate is not more paternalistic than the confusing old Food Pyramid.²⁰¹ The main difference is that it is clearer. Nor is there any serious problem with efforts to provide people with relevant facts; consider the redesigned fuel-economy label.²⁰²

2. *Disclosure: beyond the facts.* We have seen that disclosure strategies could attempt to persuade, not merely to inform. Educational campaigns involving distracted driving, seatbelt buckling, and drunk driving are examples. Graphic health warnings for cigarettes fall in the same category.²⁰³ Efforts to persuade do raise distinctive issues, and may be counted as forms of soft paternalism. But so long as freedom of choice is maintained, and deception is avoided, is it really objectionable for government to try to persuade people not to engage in behavior that causes palpable harm? The most plausible argument in favor of a negative answer is rule-consequentialist, but it is not easy to defend the assumptions on which that response must rely.²⁰⁴

3. *Default rules.* In the absence of a system of active choosing, some rule has to specify what happens if people do nothing. If, for example, it is presumed that people are not enrolled in savings plans or in health insurance programs, it is because a particular default rule has been chosen, not because God or nature

201. See Sunstein, *Empirically Informed*, *supra* note 3, at 1378-79.

202. *Id.* at 1371-76. Recall, however, that disclosure policies may affect ends as well as means. See *supra* Subsection II.A.2.

203. Sunstein, *Empirically Informed*, *supra* note 3, at 1381.

204. Cf. Glaeser, *supra* note 28 (outlining a series of objections to soft paternalism).

has so decreed. Perhaps an opt-in program is better, for various reasons, than an opt-out one; but in either case, people are being defaulted into one set of outcomes rather than another. As I have emphasized, active choosing is a possible way out of the occasional difficulty of choosing the right default rule, and in some cases, active choosing is best. But as noted above, some people would prefer not to choose, and active choosing has difficulties of its own, especially in complex or novel circumstances.²⁰⁵

4. *Fuel economy and energy efficiency.* We have seen that important federal regulations require increases in the fuel economy of motor vehicles.²⁰⁶ Related regulations require increases in the energy efficiency of household appliances, such as refrigerators.²⁰⁷ What is the rationale for such regulations? Insofar as we are dealing with air pollution, the justification is quite standard: the goal is to reduce externalities. And insofar as the goal is to promote energy security by reducing dependence on foreign oil, the justification also involves externalities. But as noted, the strong majority of the benefits from such rules do not involve air pollution or energy security. They involve consumer savings, in the form of reduced costs from use.

Should those savings be counted? On one view, they should not be. There is a market for fuel economy and for energy efficiency. Consumers can trade off the relevant values as they see fit. They can certainly purchase cars with excellent fuel economy. If they are willing to spend more on the initial purchase in order to save on gas prices over time, the market allows them ample opportunity. And if they want more fuel economy than the market now provides, they can push the market in that direction. Markets, no less than government, can be technology-forcing.

On the other hand, consumer savings from fuel-efficient cars and from energy-efficient appliances unquestionably count as benefits. They represent savings to consumers, brought about by regulation. The hard question is not deciding whether they count as benefits, but identifying the relevant market failure. A behavioral account, offered by the Department of Transportation and the EPA, emphasizes both myopia and salience. The argument here is grounded in the idea of an energy paradox, in which consumers do not purchase energy-efficient goods even though it is in their economic interest to

205. A detailed discussion is found in Sunstein, *supra* note 70, at 20-21.

206. See *CAFE—Fuel Economy*, NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., <http://www.nhtsa.gov/fuel-economy> (last visited Jan. 28, 2013).

207. See *Building Technologies Program: Appliance and Equipment Standards*, U.S. DEP'T OF ENERGY, https://www1.eere.energy.gov/buildings/appliance_standards (last updated Dec. 27, 2012).

do so.²⁰⁸ Insofar as myopia and salience are the foundations of policy, a form of behaviorally informed paternalism is involved, and it is not soft.

V. PROBLEMS WITH SOFT PATERNALISM: IMAGINARY AND REAL

The argument thus far has proceeded on the assumption that hard paternalism raises special problems and that soft paternalism is better along important dimensions. If paternalistic approaches impose small costs, or no material costs, on those who seek to go their own way, then such approaches are less vulnerable to the objections I have raised here. On the other hand, soft paternalism potentially raises three special concerns, and they should be addressed independently. The first involves transparency; the second involves the risk of manipulation; the third, which may also apply to hard paternalism, involves the legitimate claims of System 1. There are also questions about the relationship between illicit motivations and soft paternalism.

A. Of Transparency and Political Safeguards

Mandates and commands are highly visible, and government is likely to be held accountable for them.²⁰⁹ If public officials require increases in fuel economy, impose new energy efficiency requirements on refrigerators, forbid people from riding motorcycles without helmets, or require them to buckle their seatbelts, nothing is mysterious, hidden, or secret. The prohibitions may or may not be acceptable, but they lack the distinctive vice of insidiousness. No one is confused or fooled. Political safeguards are triggered. The government must defend itself publicly. And if the public defense is perceived as weak, the proposed action may well crumble. In a democracy, officials are subject to scrutiny for mandates and bans.

It is important to be careful with this argument. One person's political safeguard will be another person's interest-group power. If a mandate is vulnerable as a matter of political reality, it may not be because "the people" are unhappy; it may be because a self-interested private group is at risk and able to block a desirable measure. If a potentially life-saving policy runs into trouble because its visibility triggers opprobrium and threats of political reprisal, we do

208. For discussion of the energy paradox, see Adam B. Jaffe & Robert N. Stavins, *The Energy Paradox and the Diffusion of Conservation Technology*, 16 RESOURCE & ENERGY ECON. 91, 92-94 (1994). For a helpful and somewhat skeptical discussion, see Hunt Allcott & Michael Greenstone, *Is There an Energy Efficiency Gap?*, 26 J. ECON. PERSP. 3 (2012). It is clear that more research is needed on this topic.

209. See REBONATO, *supra* note 28, at 103-05; Glaeser, *supra* note 28, at 156.

not know that the policy is necessarily bad, and we should not take the trouble as proof of its badness. Even in a well-functioning democracy, it is important not to be naïve about the world of political safeguards; recall the public choice problem. Nonetheless, it is true that government must be accountable to the public, and visibility is, in general, an important and desirable safeguard.

This point is closely related to one made by Justice Jackson.²¹⁰ Explaining the importance of requiring the laws to be applied generally, he famously wrote:

I regard it as a salutary doctrine that cities, states and the Federal Government must exercise their powers so as not to discriminate between their inhabitants except upon some reasonable differentiation fairly related to the object of regulation. This equality is not merely abstract justice. The framers of the Constitution knew, and we should not forget today, that there is no more effective practical guaranty against arbitrary and unreasonable government than to require that the principles of law which officials would impose upon a minority must be imposed generally. Conversely, nothing opens the door to arbitrary action so effectively as to allow those officials to pick and choose only a few to whom they will apply legislation and thus to escape the political retribution that might be visited upon them if larger numbers were affected. Courts can take no better measure to assure that laws will be just than to require that laws be equal in operation.²¹¹

Justice Jackson's argument, in short, is that generality is a corrective against abuse because laws that apply generally trigger political safeguards. Friedrich Hayek wrote in similar terms, arguing that selective impositions—on religious minorities, for example—are especially troublesome because officials can impose them without fearing political retribution.²¹² By contrast, general impositions trigger the relevant safeguards and hence are less likely to be put in place unless they are justified. Indeed, Hayek's conception of the rule of law itself relies heavily on this point.

While recognizing that the existence of interest groups weakens the force of such arguments, we should be able to see that a similar point can be made about the transparency of mandates and bans, which also trigger political safeguards. On this count, some people think that soft paternalism

210. *See* *Ry. Express Agency, Inc. v. New York*, 336 U.S. 106, 112-13 (1949) (Jackson, J., concurring).

211. *Id.*

212. F.A. HAYEK, *THE CONSTITUTION OF LIBERTY* 155 (1960).

may not fare so well.²¹³ Precisely because of its subtlety and softness, such paternalism—and some nudges in particular—may be invisible. They may be manipulative.²¹⁴ They may even be self-insulating insofar as they alter the very behavior, and perhaps the public beliefs and understandings, that would otherwise be brought to bear against them. And indeed, prominent critics of soft paternalism have suggested that the lack of transparency is a serious problem. In the words of the economist Edward Glaeser:

Hard paternalism generally involves measurable instruments. The public can observe the size of sin taxes and voters can tell that certain activities have been outlawed. Rules can be set in advance about how far governments can go in pursuing their policies of hard paternalism. Effective soft paternalism must be situation specific and creative in the language of its message. This fact makes soft paternalism intrinsically difficult to control and means that it is, at least on these grounds, more subject to abuse than hard paternalism.²¹⁵

The underlying concern must be taken seriously, and the best response is simple. For Glaeser's reasons, nothing should be hidden, and everything should be transparent. Soft paternalism, nudges, and any other behaviorally informed approaches, no less than hard paternalism, should be visible, scrutinized, and monitored. To the extent feasible, rules that embody soft paternalism should be subject to public scrutiny in advance, often through notice-and-comment rulemaking. Consider some initiatives: automatic enrollment in savings and health care plans; the substitution of the Food Plate for the Food Pyramid and other disclosure policies that reflect how people actually process information; graphic health warnings; fuel-economy standards. All of these initiatives are visible, public, and entirely observable. All were, and remain, subject to public scrutiny. None is "intrinsically difficult to control." In this light, what is the problem?

Perhaps we can answer by offering a behavioral twist on Glaeser's argument. Perhaps the problem is not so much a lack of transparency as a lack of salience. Perhaps the objection is not that the government keeps its initiatives secret, but that they do not attract the kinds of attention that are typically triggered by mandates and bans.

So understood, the objection is certainly plausible insofar as it applies to actions of government that, while hardly hidden, lack the kinds of salience that

213. See Glaeser, *supra* note 28, at 149.

214. See CONLY, *supra* note 28, at 8, for the suggestion that nudging can be both manipulative and ineffective, thus giving us "the worst of both worlds."

215. See Glaeser, *supra* note 28, at 151.

produce careful public scrutiny. Salience matters for laws and regulation as for everything else. The problem is that nudges and soft paternalism are often highly salient, and the line between what is salient and what is not is hardly the same as the line between hard paternalism and soft paternalism. Consider, for example, graphic warnings for cigarettes, which received considerable public scrutiny and attention, and the new fuel-economy labels, which were also highly salient to the public. A lack of salience does create a problem for those who emphasize political safeguards, but soft paternalism may be highly salient; there is nothing intrinsic to nudging that reduces salience.

B. Of Easy Reversibility

In imposing very low costs or in failing to impose material costs on choices, soft paternalism differs from mandates and bans. Because of the absence of such costs, soft paternalism appears to be *easily reversible*.

For example, graphic warnings do not override individual choice, and while they are not neutral and are meant to steer, people can ignore them if they want. We can easily imagine, and even find, graphic warnings that are meant to discourage texting while driving, abortion, premarital sex, and gambling. However powerful, any such warnings can be ignored. Those who run cafeterias and grocery stores might place fruits and vegetables at the front and cigarettes and fatty foods at the back. Even if so, people can always go to the back. A default rule in favor of automatic enrollment—in a savings or health insurance plan or a privacy policy—will greatly affect outcomes, and may be decisive for many of us. But we can always opt out.

Does this mean that so long as soft paternalism or a nudge is involved, no one should worry about paternalism, or indeed about any abuse of authority or power? That would be an unwarranted and potentially dangerous conclusion. We can readily identify an important problem with the idea of easy reversibility: the very biases and decisional inadequacies that I have traced here suggest that even when reversibility is easy in theory, it may prove difficult in practice. In part because of the power of System 1, soft paternalism may be decisive.

True, we can search for chocolate candy and cigarettes at the back of the store, and true, we might opt out of a website policy that authorizes a lot of tracking (perhaps with a simple click)—but because of the power of inertia, many of us are not likely to do so. Graphic warnings, which appeal directly to System 1, may be exceedingly effective precisely because they target identifiable features of human cognition. The idea of easy reversibility might, in these circumstances, seem a bit of rhetoric, even a fraud—perhaps comforting, to be sure, but not a realistic response to those who are concerned about potential

errors or bad faith on the part of soft paternalists and nudgers.

This objection has unquestionable force. It would be misleading to suggest that because of easy reversibility, all risks are eliminated. If people are defaulted into exploitative savings plans (with high fees and little diversification) or unduly expensive health insurance programs, it is not enough to say that they can go their own way if they choose to do so. If a website allows you to opt out of a privacy policy that allows it to track all of your movements on the Internet, you may say, “yeah, whatever,” and not alter the default. Magazine subscribers who no longer enjoy the magazines to which they subscribe, but whose subscriptions are automatically renewed, often do not take the trouble to discontinue them. (How often do you click on the “unsubscribe” button when you find yourself on an unwanted email list?)

In view of the fact that people do not opt out even when it is easy to do so, a self-interested or malevolent government could easily use soft paternalism to move people in its preferred directions. If we accept very strong assumptions about the likelihood of government mistake and about the virtues of private choice (uninfluenced by government), we might reject soft paternalism, at least where it is not inevitable.

It remains true, however, that insofar as it maintains freedom of choice, soft paternalism is less intrusive and less dangerous than mandates and bans. This is so even if people will exercise that freedom less often than they would if inertia and procrastination were not powerful forces. It is important to emphasize that in the face of bad defaults, a number of people will in fact opt out.²¹⁶ If people are defaulted into a retirement plan that puts a lot of their money into savings while giving them too little now, they will indeed reject the default.²¹⁷ If people are defaulted into a health insurance plan that works out very badly for them, many of them are going to switch. For that reason, liberty of choice is a real safeguard. We have seen enough to know that the freedom to opt out is no panacea. But it is exceedingly important.

C. The Legitimate Claims of System 1

Here is one way to understand one of the claims made here. Because of System 1, people err. We need to strengthen the hand of System 2 by promoting self-control, unshrouding attributes, counteracting biases, and eliminating an undue focus on the short-term. Some forms of paternalism

216. See, e.g., Sunstein, *supra* note 70, at 16-17.

217. See John Beshears et al., *The Limitations of Defaults 8* (Sept. 15, 2010) (unpublished manuscript), <http://www.nber.org/programs/ag/rrc/NB10-02,%20Beshears,%20Choi,%20Laibson,%20Madrian.pdf>.

move people in the directions that they would go if they were fully rational. Paternalism, whether hard or soft, creates “as-if” rationality. Indeed, that is a central point of good choice architecture.

It would be possible to object that if this approach is understood in a certain way, it ignores the legitimate claims of System 1.²¹⁸ More bluntly, it disregards a lot of what is most important in and to human life. Some of our favorite foods are pretty fattening. To be sure, most people care about their health, but unless they are fanatical, their health is hardly the only thing they care about. Many people like to drink and to smoke. Many of us care more about current consumption than consumption twenty years from now. When people enjoy their lives, it is because of System 1. The future matters, but the present matters too, and people reasonably and legitimately strike their own balance. Why should public officials, or anyone, make people focus on something other than what they want to focus on, and promote choice architecture that devalues, denigrates, and undermines some of their most fundamental motivations and concerns?²¹⁹ Indeed, might not System 2 be paralyzed if it lacks a sense of those concerns? How will it know what to do?

Consider a patient of Antonio Damasio, who suffered from brain damage that prevented him from experiencing emotions.²²⁰ Because the patient lacked “gut reactions,” he could perform some tasks well; for example, he was able to drive safely on icy roads, avoiding the natural reaction to hit the brakes during a skid. On the other hand, his ability to focus on consequences was accompanied by extreme difficulty in making decisions:

I was discussing with the same patient when his next visit to the laboratory should take place. I suggested two alternative dates, both in the coming month and just a few days apart from each other. The patient pulled out his appointment book and began consulting the calendar. . . . For the better part of a half-hour, the patient enumerated reasons for and against each of the two dates: previous engagements, proximity to other engagements, possible meteorological conditions, virtually anything that one could reasonably think about concerning a single date. Just as calmly as he had driven over the ice, and recounted that episode, he was now walking us through a tiresome cost-benefit analysis, an endless outlining and fruitless comparison of options and possible consequences.

218. See REBONATO, *supra* note 28.

219. This question is pressed in *id.*; and Wright & Ginsburg, *supra* note 28.

220. ANTONIO DAMASIO, DESCARTES’ ERROR: EMOTION, REASON AND THE HUMAN BRAIN 192-93 (2006).

....

[We] finally did tell him, quietly, that he should come on the second of the alternative dates. His response was equally calm and prompt. He simply said: "That's fine."²²¹

Without emotions, and without System 1, might we not be endlessly listing consequences while lacking a motivation for choosing among them?

The underlying questions are legitimate, and they suggest the problems with some imaginable nudges. At the same time, they reflect a misunderstanding of the argument made here, and the approaches that it supports. To see why, we need to make a distinction between two different understandings of the kinds of biases and errors that paternalism might counteract.

The most ambitious view might be associated with Mr. Spock of the old *Star Trek* show and the new *Star Trek* movies. (Aficionados might consider this the Vulcan view, after Mr. Spock's logic-dominated planet.) On that view, an understanding of bounded rationality and of cognitive biases suggests that System 2 needs to be put firmly in charge. To the extent that it is not, choice architecture should be established to move people to a situation of "as-if" System 2 primacy. This view raises many puzzles, because some of the greatest pleasures of life appeal directly to System 1. No sensible private or public institution would be indifferent to the fact that for sufficient reasons, people take risks because that is what they like to do. They fall in love; they overeat; they stay up all night; they get drunk; they act on impulse; they run with apparently unpromising ideas; they experiment in a million and one different ways.

On a less ambitious view, bounded rationality and cognitive biases lead people to make what they themselves see as serious errors, or would see as serious errors after reflection, and choice architecture should be established to help make those errors less likely or less damaging. If inertia leads people not to take action that (they do or would agree) is in their interest, then inertia might be enlisted to promote outcomes that (they do or would agree) are in their interest. If a problem of self-control is leading people to endanger their health, and if they do or would want private and public institutions to help to solve that problem (and not to exploit it), then there is no cause for complaint if they do so. We need not denigrate the legitimate claims of System 1 in order to accept these points. The real problem lies not in any question of high principle, but in identifying what people do or would want, and in deciding whether choice architects can be trusted.

221. *Id.* at 193-94.

With respect to issues of this kind, there are limits to how much progress can be made in the abstract. We need to ask concrete questions about concrete problems. We could imagine forms of paternalism that would be objectionable because they would neglect what people really care about. Consider the “Suffer Now, Celebrate When You’re Almost Dead Pension Plan,” automatically putting fifty-one percent of employee salaries into savings; or the “You Will Hate It Wellness Program,” asking employees to commit to a grueling and unpleasant daily exercise regime; or the “Joyless Cafeteria,” keeping the tastiest foods relatively hidden. We could also imagine paternalistic approaches that are helpful rather than harmful. The challenge is to avoid the latter and promote the former.

D. A Real Concern: Impermissible Motivations

I have emphasized that soft paternalism does not impose material costs on choices. Even so, it is correct to object that this point is not conclusive and that some forms of soft paternalism would go beyond the appropriate line. The most problematic cases reflect not unacceptable paternalism but a different problem: *impermissible motivations*. Indeed, many of the strongest intuitive objections to paternalism, even in its soft form, involve examples, real or imagined, in which government is acting on the basis of impermissible factors. The objections are right, but the real problem has nothing to do with paternalism.²²²

We would not, for example, want to authorize government to default people into voting for incumbents by saying that unless they explicitly indicate otherwise, or actually show up at the ballot booth, they are presumed to vote for incumbents. Or suppose it were declared that for purposes of the census, citizens are presumed to be Caucasian, unless they explicitly state otherwise. Some information campaigns are unacceptable for the same reason. Suppose that government decided to inform people about all the misdeeds committed by members of a particular religious faith (say, Catholics or Jews). Or suppose that government decided to use vivid images to convince people to choose products manufactured by its favorite interest groups.

In all of these cases, the problem does not lie with paternalism. The

222. See CONLY, *supra* note 28, at 103-12, for a discussion of paternalism and perfectionism. Conly supports hard paternalism, but only as a means of promoting people’s own ends, and not in order to displace them. She rejects perfectionism. The paternalism defended here is similarly focused on means, not ends (with the qualifications I have explored); my discussion is thus compatible with Conly’s rejection of perfectionism. It would be possible to go further than the treatment in this section and to say, with Conly, that any form of paternalism, soft or hard, is unacceptable if it is perfectionist in character.

problem is the illegitimate or illicit ends that official paternalism, even if soft, is meant to produce. In a free and democratic society, government is not supposed to use the basic rules of voting to entrench itself, to favor certain racial groups, to stigmatize members of a particular faith, or to tell people to buy the products that its favored interest groups manufacture. When the government's ends are illicit, paternalism, designed to promote those ends, is illicit too.

We can imagine cases in which the illicit nature of the government's ends is clear; the examples given above are meant to be such cases. But we can also imagine cases about which people might disagree. Suppose, for example, that government were to engage in soft paternalism—say, through an educational campaign—designed to discourage people from having sex before marriage or from choosing abortion. Some people might think that efforts of this kind would be illicit, because they would violate a commitment to neutrality in the relevant domains. Perhaps those people are right; perhaps not. In either case, the central question would be whether the government's ends were illicit; it is not about paternalism.

The examples of illicit ends are important because they place some limits on even minimally intrusive forms of paternalism. But with respect to the issues under discussion here, they are uninformative, because they do not establish the central claim, which is that certain forms of paternalism are objectionable as such. If some people are strongly committed to that claim, it is not clear what might be said to dislodge that commitment. Is it really an insult to autonomy to provide graphic images of the harms associated with cigarette smoking? To undertake an educational campaign in favor of healthy eating or against texting while driving? For those who think so, the risk is that high-sounding abstractions are being enlisted to prevent initiatives that insult no one and that promise to make people's lives healthier, longer, and better.

CONCLUSION

My goal here has been to explore the relationship between human error and paternalism. I have urged that accumulating evidence suggests, more concretely than ever before, that in identifiable cases, people's choices can produce serious harm, even when third parties are not at risk. The result is a series of behavioral market failures that provide at least a plausible basis for some kind of official remedy. We have identified a general law of behaviorally informed regulation, which is that the appropriate responses to behavioral market failures generally consist of nudges, usually in the form of disclosure,

warnings, and default rules.²²³ These responses do not attempt to revisit people's ends.²²⁴ They are focused on correcting mistakes that people make in choosing the means that would promote their own ends.

On grounds of autonomy, some forms of paternalism seem objectionable. I have raised the possibility that objections from autonomy may, in this context, be a heuristic for what really matters, which is welfare. Even if this argument is unconvincing—and I have done little more than to gesture toward it here—the most sensible responses to behavioral market failures do not run afoul of autonomy, properly conceived, and certainly not if they take the form of nudges.

The most forceful objections are welfarist in character. For a variety of reasons, the cure may be worse than the disease. Indeed, the disease itself may produce long-term benefits, not least in the form of learning. At the same time, we have seen that choice architecture is inevitable, whether or not it is intentional or a product of any kind of conscious design. We have also seen that the strongest objections to hard paternalism are weaker when applied to soft paternalism.

In many cases, the welfarist arguments against paternalism, whether hard or soft, have considerable force; but they depend on normative claims that are complex and contested, and on empirical claims that may not be true. We do best to avoid high-sounding abstractions and general propositions that are merely plausible. A central question is the costs and the benefits of particular approaches, whether paternalistic or not.²²⁵ Understandings of behavioral market failures, and of the promise of choice architecture, are uncovering many opportunities for increasing people's welfare without compromising the legitimate claims of freedom of choice. We will uncover many more such opportunities in the future.

Let's take advantage of them.

223. I have also noted that in some cases, a stronger response may be justified after careful consideration of benefits and costs. To the same effect, see *id.* at 28.

224. Recall that this proposition must be qualified. See *supra* Subsection II.A.2.

225. It is true that we can imagine cases in which the quantifiable benefits of certain approaches justify the quantifiable costs, but in which some other value, such as dignity or privacy, imposes a constraint. See SUNSTEIN, SIMPLER, *supra* note 3, for relevant discussion.