NELLCO
NELLCO Legal Scholarship Repository

New York University Public Law and Legal Theory Working Papers

New York University School of Law

4-16-2008

Little Brother is Watching You: New Paternalism on the Slippery Slopes

Mario J. Rizzo
Department of Economics, NYU, mario.rizzo@nyu.edu

Douglas G. Whitman
CSUN, glen.whitman@csun.edu

Follow this and additional works at: http://lsr.nellco.org/nyu_plltwp

Part of the Jurisprudence Commons, and the Law and Economics Commons

Recommended Citation
Rizzo, Mario J. and Whitman, Douglas G., "Little Brother is Watching You: New Paternalism on the Slippery Slopes" (2008). *New York University Public Law and Legal Theory Working Papers.* Paper 71.
http://lsr.nellco.org/nyu_plltwp/71

This Article is brought to you for free and open access by the New York University School of Law at NELLCO Legal Scholarship Repository. It has been accepted for inclusion in New York University Public Law and Legal Theory Working Papers by an authorized administrator of NELLCO Legal Scholarship Repository. For more information, please contact tracy.thompson@nellco.org.

02/27/09

LITTLE BROTHER IS WATCHING YOU
New Paternalism on the Slippery Slopes

By Mario J. Rizzo[*] and Douglas Glen Whitman[**]

"The Child is father of the Man."
William Wordsworth

ABSTRACT: The "new paternalism" claims that careful policy interventions can help people make better decisions in terms of their own welfare, with only mild or nonexistent infringement of personal autonomy and choice. This claim to moderation is not sustainable. Applying the insights of the modern literature on slippery slopes to new paternalist policies suggests that such policies are particularly vulnerable to expansion. This is true even if policymakers are fully rational. More importantly, the slippery-slope potential is especially great if policymakers are not fully rational, but instead share the behavioral and cognitive biases attributed to the people their policies are supposed to help. Accepting the new paternalist approach creates a risk of accepting, in the long run, greater restrictions on individual autonomy than have been heretofore acknowledged.

OUTLINE:

1. Introduction
2. An Introduction to the New Paternalism
3. An Introduction to Slippery Slopes
4. Gradients and Vagueness in the New Paternalism
    a. How Gradients Encourage Slippery Slopes
    b. How the New Paternalism Creates New Gradients
    c. How the New Paternalism Exploits Existing Gradients
        i. Hyperbolic and Quasi-hyperbolic Discounting
        ii. Framing and Context-Dependence
5. Slippery Slopes with Rational Policymakers
    a. Altered Incentives Slopes
        i. Bias Interactions
        ii. Crowding Out of Self-Regulation
        iii. Substitutability of Personal Inputs
    b. Authority, Simplification, and Distortion Slopes
        i. Simplification of Theory by Experts
        ii. Distortion of Facts by Rent-Seekers
    c. Expanding Justification Slopes
        i. The Logic of Justificatory Expansion

[*] Department of Economics, New York University.
[**] Department of Economics, California State University, Northridge.

ii. Application to Smoking Bans
d. On Experts vs. Regular People
6. Slippery Slopes with Cognitively Biased Policymakers
a. Hyperbolic Discounting by Policymakers
b. Narrow Framing by Policymakers
c. Passive Framing Accepted by Policymakers
d. Extremeness Aversion
e. Prototype and Affect Heuristics: Extension Neglect
i. Extension Neglect in the New Paternalist Framework
ii. Extension Neglect in the Calculation of Optimal Sin Taxes
7. Rejoinder and Recommendations
a. Rejoinder to New Paternalist Responses
b. Minimizing the Danger

1. INTRODUCTION

Paternalist arguments advocate forcing or manipulating individuals to change their behavior for their own good, as distinct from the good of others. Paternalism has been with us for millennia. Recently, however, a seemingly new form has arisen that we call "the new paternalism."[1] Unlike the old paternalism, which sought to make individuals behave consistently with the (often moralistic or religious) preferences of policymakers, the new paternalism seeks to help individuals maximize their own welfare as they see it themselves (at some deep level). Richard Thaler and Cass Sunstein's recent book *Nudge: Improving Decisions About Health, Wealth, and Happiness*[2], one of *The Economist*'s books of the year for 2008,is an important summary statement of the new paternalist paradigm – as is Daniel Ariely's *Predictably Irrational: The Hidden Forces that Shape Our Decisions*.[3]

The new paternalism is supported by a growing body of research in behavioral economics showing that individuals are not fully "rational," as economists understand that term, but instead are subject to a variety of cognitive errors and biases. The list of such deviations from strict rationality includes – but is not limited to – status quo bias[4], optimism bias[5],

---

[1] Colin Camerer et al., *Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism,'* 151 U. PA. L. REV. 1211 (2003); Richard H. Thaler & Cass R. Sunstein, *Libertarian Paternalism*, 93 AEA PAPERS AND PROCEEDINGS 175 (2003); Cass R. Sunstein and Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159 (2003); Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, 35 J. LEGAL STUD. 199 (2006); Jonathan Gruber & Botond Koszegi, *Is Addiction 'Rational'? Theory and Evidence*. 116 Q. J. ECON. 1261 (2001); Ted O'Donoghue & Matthew Rabin, Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes, 93 AEA PAPERS & PROCEEDINGS 186 (2003); Ted O'Donoghue & Matthew Rabin, *Optimal Sin Taxes* (2003) (unpublished manuscript, on file with Cornell University, University of California at Berkeley).
[2] Richard H. Thaler & Cass R. Sunstein, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008).
[3] Daniel Ariely, PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS (2008).
[4] William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988).
[5] W. Kip Viscusi & Wesley A. Magat, LEARNING ABOUT RISK: CONSUMER AND WORKER RESPONSE TO HAZARD INFORMATION, Cambridge, Mass: Harvard University Press (1987), 93-95.

susceptibility to framing effects[6], and lack of willpower or self-control.[7] Thus individuals are viewed as "pawns in a game whose forces [they] largely fail to comprehend."[8] To the extent that these cognitive problems cause individuals to make systematic and predictable choices inconsistent with their own well-considered preferences, there is potentially room for paternalistic interventions that will help them do better.[9] In fact, these interventions have been describes as "free lunches…that would help people achieve more of what they *truly want*."[10]

The use of behavioral economics to justify paternalism has been criticized on various grounds. At the most philosophical level, there are serious questions about the underlying values and welfare standards implicit in the new paternalism.[11] At the policy level, we have argued elsewhere that policymakers may lack the knowledge necessary to craft beneficial paternalist policies.[12] The policies in question could produce ineffective or even counterproductive results, because they interfere with individuals' self-debiasing and learning processes.[13] Policymakers may also lack the proper incentives to implement wise policies, given their own self-interest and the lobbying efforts of interested parties.[14] If policymakers are subject to the same cognitive biases that afflict regular people, that, too, will inhibit good policymaking.[15]

In this article we address another aspect of the new paternalism: its vulnerability to slippery slopes that can lead from modest (or "soft") paternalism to more extensive (or "hard") paternalism.

New paternalists distance themselves from hard paternalism by emphasizing the moderate character of their proposals. Christine Jolls and Cass Sunstein frequently refer

---

[6] Colin F. Camerer, *Prospect Theory in the Wild: Evidence from the Field*, in Daniel Kahneman & Amos Tversky, eds, CHOICES, VALUES, AND FRAMES, Cambridge (2000), 288, 298-295.
[7] Richard H. Strotz, *Myopia and Inconsistency in Dynamic Utility Maximization*, 23 REV. ECON. STUD. 165 (1955/56); Shane Frederick, George Loewenstein, & Ted O'Donoghue, *Time Discounting and Time Preference: A Critical Review*, 40 J. ECON. LIT 351 (2002); George Ainslie, BREAKDOWN OF WILL, Cambridge University Press (2001).
[8] Ariely (2008), 243.
[9] For a fuller description of the policies in question, see Part 2.
[10] Ariely (2008), 241-2, emphasis added.
[11] Gregory Mitchell, *Libertarian Paternalism Is an Oxymoron*, 99 NORTHWESTERN UNIVERSITY L. REV. 1245 (2005); Mario J. Rizzo & Douglas Glen Whitman, *Meet the New Boss, Same as the Old Boss: A Critique of the New Paternalism* (2007) (unpublished draft manuscript, on file with New York University).
[12] Mario J. Rizzo & Douglas Glen Whitman, *The Knowledge Problem of New Paternalism* , BRIGHAM YOUNG UNIVERSITY LAW REVIEW (forthcoming, 2009).
[13] Jonathan Klick & Gregory Mitchell, *Government Regulation of Irrationality: Moral and Cognitive Hazards*, 90 MINN. L. REV. 1620 (2006); Glen Whitman, *Against the New Paternalism: Internalities and the Economics of Self-Control*, CATO INSTITUTE POLICY ANALYSIS no. 563 (2006).
[14] Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133 (2006).
[15] Glaeser (2006). For more informal critiques of the new paternalism (especially Sunstein & Thaler's "libertarian paternalism"), see the weblog posts: Richard Posner, *Libertarian Paternalism – Posner's Comment*, The Becker-Posner Blog (Jan. 14, 2007), available at http://www.becker-posner-blog.com/archives/2007/01/libertarian_pat.html; and Gary Becker, Libertarian Paternalism: A Critique, The Becker-Posner Blog (Jan. 14, 2007), available at http://www.becker-posner-blog.com/archives/2007/01/libertarian_pat_1.html.

to their proposals for debiasing behavior through law as a "middle ground" between laissez-faire and more heavy-handed paternalism[16], one that is a "less intrusive, more direct, and more democratic response to the problem of bounded rationality."[17] Colin Camerer, et al., present their model of "asymmetric paternalism" as "a careful, cautious, and disciplined approach" to evaluating paternalistic policies.[18] Cass Sunstein and Richard Thaler characterize their "libertarian paternalist" approach as a "relatively weak and nonintrusive type of paternalism" that in its "most cautious forms … imposes trivial costs on those who seek to depart from the planner's preferred option."[19] In short, the new paternalists claim we can attain significant improvements in individual welfare with relatively small interventions that do not substantially restrict liberty or autonomy.

Our thesis is that the new paternalism's claim to moderation is not sustainable. A recent literature, to which we have contributed, has rehabilitated slippery-slope reasoning by examining the specific processes by which slippery slopes occur, as well as the circumstances under which slippage is most likely.[20] Applying the insights of the slippery-slope literature to new paternalist policies suggests that these policies are particularly subject to expansion. We argue that is true even if policymakers are rational. But perhaps more importantly, we argue that the slippery slope threat is especially great if policymakers are not fully rational, but instead share the behavioral and cognitive biases attributed to the people their policies are supposed to help. Consequently, accepting new paternalist policies creates a risk of accepting, in the long run, greater restrictions on individual autonomy than have been heretofore acknowledged. Inasmuch as new paternalists claim to be interested in preserving autonomy[21], this surely must be counted as an unrecognized or unacknowledged cost to be balanced against any possible gains from their policies.

New paternalists are not entirely unaware of this possibility. In fact, they have invited scholars to explore the slippery slope potential of their policy suggestions.[22] We accept the invitation.

---

[16] Jolls & Sunstein 2006, 208, 216.
[17] Ibid., 201.
[18] Camerer, et al., 2003, 1212.
[19] Sunstein & Thaler (2003), 1162.
[20] Douglas Walton, SLIPPERY SLOPE ARGUMENTS (1992); Sanford Ikeda, DYNAMICS OF THE MIXED ECONOMY: TOWARD A THEORY OF INTERVENTIONISM (1997); Eugene Volokh, *The Mechanisms of the Slippery Slope*, 116 HARV. L. REV. 1026 (2003); Mario J. Rizzo and Douglas Glen Whitman, *The Camel's Nose Is in the Tent: Rules, Theories, and Slippery Slopes*, 41 UCLA L. REV. 539 (2003), Eric Lode, *Slippery Slope Arguments and Legal Reasoning*, 87 CALIF. L. REV. 1469 (1999); Frederick Schauer, *Slippery Slopes*, 99 HARV. L. REV. 361 (1985).
[21] E.g., Sunstein & Thaler (2003), 1161: "To borrow a phrase, libertarian paternalists urge that people should be 'free to choose.' Hence we do not aim to defend any approach that blocks individual choices." (Internal footnote omitted.)
[22] E.g., Camerer et al. (2003), 1251: "The potential for such 'slippery slopes' commonly arises in policy debates and clearly arises here as well. But just as for other domains, the ideal way to deal with these possibilities is not to avoid policy changes altogether, but to consider the extent to which future policies are made to appear more or less attractive by the one under consideration."

In sections 2 and 3, we clarify the meaning of new paternalism and slippery slopes, respectively. In section 4 we discuss the close relationship between slippery slopes and gradients – a theme that will recur through the remainder of the article. In section 5, we consider paternalist slopes that can occur even if policymakers are fully rational. In section 6, we discuss how paternalist slopes become even more likely if policymakers are subject to the same cognitive and behavioral biases that are presumed to affect regular people. Finally, in section 7, we offer both rejoinders to arguments against slippery-slope analysis and recommendations for how to resist paternalist slopes.

2. AN INTRODUCTION TO THE NEW PATERNALISM

The new paternalists are partially wedded to the principle of standard economics that an agent's welfare ought to be defined in terms of the goals or purposes of the agent himself. Unlike the standard welfare economist, however, the new paternalist is not satisfied with the overt expressions of the agent. Real people display cognitive and behavioral biases, including "self-control problems," "failure to process information as Bayes's rule would require," and "systematic mispredictions about the costs and benefits of choices."[23] As a result, the "revealed preferences" or actual choices of the individual are often a poor guide to his well-being. The individual's true well-being must therefore be discovered through other means, such as psychological experiments, responses to surveys, or actual choices made under ideal conditions of perfect information and sober reflection.

As noted earlier, the new paternalists claim to be moderate; that is, they advocate policies that interfere with individual choice or autonomy only slightly and yet promise substantial welfare gains. The two leading statements of the new paternalism are Camerer and coauthors' "asymmetric paternalism" and Sunstein and Thaler's "libertarian paternalism." The former refers to policies that create "large benefits to those who are boundedly rational… while imposing little or no harm on those who are fully rational."[24] The latter describes just how this is possible: "The libertarian paternalist insists on preserving choice, whereas the non-libertarian paternalist is willing to foreclose choice."[25] In other words, the libertarian paternalist will agree to policies that *frame* or structure people's decisions in a certain way – for example, by requiring that employees be automatically enrolled in retirement plans unless they specifically opt out, or by automatically granting employees a right not to be dismissed except "for cause" unless they voluntarily waive it.

Notwithstanding their stated concern for individual autonomy, the new paternalists are willing to accept more coercive policies if the policymaker's degree of assurance that his policy will benefit agents is high – a point on which we will expand later. Policies that interfere slightly with individual choice are often called "soft" while those that interfere more are called "hard." Framing decisions by compelling employers to automatically enroll workers in savings plans is soft paternalism; banning trans-fats outright is hard. A

[23] Camerer et al. (2003), 1217-1218.
[24] Camerer et al. (2003), 1219.
[25] Sunstein & Thaler (2003), 1185.

variety of other policies, such as sin taxes, mandatory contractual terms, and cooling-off periods, would fall somewhere in between.

The set of policies that might be considered "new paternalist" is very broad, and its boundaries not entirely clear. A more complete list of policies, roughly in order of increasing intrusiveness, is presented in section 4b.

3. AN INTRODUCTION TO SLIPPERY SLOPES

The term "slippery slopes" is shorthand for two related phenomena: slippery slope *arguments* and slippery slope *events*. A slippery slope argument (SSA) is an argument about how the acceptance of one argument (regarding a decision, act, or policy) may lead to the acceptance of other arguments (regarding other decisions, acts, or policies). It describes "a 'process' or 'mechanism' by which accepting the initial argument and making the initial decision raise the likelihood of accepting the later argument and making the later decision."[26]

Importantly, SSAs do not describe inevitabilities, but simply tendencies or increases in the probability of unfavorable outcomes. A slippery slope event (SSE) refers to the actual manifestation of the events (decisions, acts, or policies) described in the SSA.[27]

SSAs and SSEs may, in principle, involve only one actor – say, a Robinson Crusoe decisionmaker. Crusoe might, for example, start by accepting an argument about the value of relaxation and end up accepting an argument in favor of serious laziness (perhaps because there is, along the way, no clear dividing line between the two).

Most slippery slopes, however, involve more than one actor. For example, if the government imposes a policy that protects people from the consequences of their mistakes (e.g., national health insurance that covers the consequences of poor health choices), it may encourage moral hazard and thus result in more mistakes (more bad health choices). Here there are at least two sets of actors involved: the policymakers who adopted the policy and the people affected by it. In turn, the behavior of the affected agents may motivate the adoption of additional policies (such as regulation of health decisions); and since the composition of legislatures changes over time, the new policymakers may differ from the initial policymakers.

That most slippery slopes involve multiple actors is a point that bears emphasis, since critiques of slippery-slope reasoning often miss it. A slippery-slope skeptic might simply say, "We'll do the right thing today, and then resist doing the wrong thing tomorrow." The problem is that the content of "we" can change. The policymakers who create the

[26] Rizzo & Whitman (2003), 544.
[27] Rizzo & Whitman (2003), 545. If an SSA provides a highly persuasive cautionary tale, it may help to avert an SSE – possibly by persuading people to oppose the initial policy. Thus, an SSA could be correct and yet the corresponding SSE is never observed.

initial policy are not necessarily the same as those who will consider a subsequent policy, nor are they coextensive with the people affected by their policies.

It is therefore useful to distinguish clearly among the various actors involved in a slope process. We concentrate on four groups: (1) *experts* such as scientists and economists; (2) *policymakers* such as politicians, bureaucrats, judges, and voters; (3) *target agents*, that is, those people whose actions are to be controlled or influenced; and finally, (4) *rent seekers*, or those who seek regulations for their own private, but not necessarily monetary, gain regardless of public interest concerns. (An example of the last group would be mutual fund companies that have an interest in encouraging more savings regardless of the targets' overall welfare or groups ideologically opposed to smoking regardless of the smokers' own preferences.)

Authors in the slippery-slope literature have also emphasized that there is no *single* slippery slope phenomenon. Instead, there are various processes, or mechanisms, by which slippery slopes can occur.[28] These processes are the key to understanding the "logic" of how one argument or policy can lead to another. The diversity of slippery slope processes will become apparent in sections 5 and 6. In section 5, we will work on the assumption that only the target agents are subject to cognitive or behavioral biases[29], while the other groups, including policymakers, are fully rational. The point is not that these slippery slopes *require* rational policymakers, but that they may occur *even with* rational policymakers. Later, in section 6, we will allow policymakers and experts to be subject to biases as well.

4. GRADIENTS AND VAGUENESS IN THE NEW PATERNALISM

a. How Gradients Encourage Slippery Slopes

As various slippery-slope analysts have recognized, slippery slopes flourish in the presence of a gradient or continuum.[30] When arguments or policies are connected by a series of small (perhaps infinitesimally small) steps, the absence of a sharp line between different cases eases the process of moving from one to another.

Gradients typically result from the vagueness of a key term. For example, there is no precise number of years that separates a "child" from an "adult;" there is no specific IQ that separates the "mentally able" from the "mentally retarded." Though we may choose an arbitrary dividing line for a particular purpose (e.g., 18 years for legal majority), there is nothing inherently right about it. People on either side of the line may be virtually indistinguishable. We call this *continuity vagueness* because it exists in the presence of a measurable variable.

---

[28] See especially Volokh (2003) and Rizzo & Whitman (2003).

[29] If targeted agents are not, in fact, subject to cognitive biases, then the case for new paternalist is already mistaken. So we take it as given, for argument's sake, that enough people have great enough cognitive biases to make new paternalist policies at least *prima facie* desirable.

[30] See especially Rizzo & Whitman (2003), 557-560; Volokh (2003), 1105-1114; Lode (1993), 1477ff.

On the other hand, *similarity vagueness* exists when measurement is not possible, irrelevant, or when it depends, at least in part, on imprecise components. We might say, for instance, that a butter knife is similar to a steak knife, a steak knife similar to a dagger, a dagger similar to a sword, and (perhaps) a sword similar to a gun. Similarity relationships are inherently vague because similarity is not precisely definable; it is intuitive and elusive.[31] A judgment of similarity can be based on both objective and subjective (or impressionistic) components along many dimensions.

The existence of a gradient created by vagueness (whether from continuity or similarity) does not necessarily lead to a slippery slope. It is sometimes possible to resist the slope, perhaps by standing firmly on an arbitrary distinction. But the existence of a gradient makes defending a given position harder than it would be otherwise, because no specific line can be defended in principle. And in the presence of similarity (rather than continuity) vagueness, even drawing an arbitrary line can prove difficult.

Movement along a gradient is especially likely in the context of precedent-based decisionmaking, as in a common law judicial system. Given that no two cases are exactly alike, precedent can operate only if decisionmakers rely on judgments of continuity or similarity. If there are relatively sharp lines between classes of case, the slippery slope threat from precedential decisionmaking is small. But in the presence of a gradient, the slippery slope threat is larger, as a sequence of "close" cases that differ only slightly can provide a bridge between cases that differ substantially.

Judicial decisionmaking is especially vulnerable to gradients, according to Eric Lode, because of judges' tendency "to place a premium both on drawing non-arbitrary, rationally defensible lines and on maintaining a coherent, consistent body of case law within a particular jurisdiction."[32] Legislators, on the other hand, can more easily impose arbitrary distinctions. But legislative decisionmaking is not immune to slipping on gradients, for at least three reasons.

First, the fact that legislators *can* impose arbitrary lines does not necessarily mean they *will*. Their political incentives can militate against taking unambiguous stands that create clear winners and losers, and they may better serve their interests by ceding discretion to bureaucrats or judges. Political scientist Gary Bryner points out that "[m]ost regulatory laws, however, give little guidance to agencies for the substance of their regulations and for the way in which the burdens they impose are to be distributed. … Some laws provide competing objectives that give administrators broad latitude."[33]

---

[31] D.H. Rouvray, *Definition and Role of Similarity Concepts in the Chemical and Physical Sciences*, 32 JOURNAL OF CHEMICAL INFORMATION AND COMPUTER SCIENCES 580 (1992), 580-86: "In spite of there being manifold examples of similarity about us, this does not mean that the concept is either easy to comprehend or to define. In fact, the concept is remarkably elusive, and it is fair to say that the concept cannot be defined in any absolute sense. However we may choose to define the similarity of two entities or events, there will always be some arbitrariness associated with whatever measure we adopt."

[32] Lode (1999), 1494.

[33] Gary C. Bryner, BUREAUCRATIC DISCRETION: LAW AND POLICY IN FEDERAL REGULATORY AGENCIES (1987), 7.

Second, legislators and bureaucrats are subject to the pressures of lobbying by special interests. Such groups may have an interest in pushing for small changes that gradually move policy along the gradient.[34] The groups in question might have a financial interest in doing so; for instance, milk producers could favor ever-greater restrictions on the availability of soft drinks, and financial services firms could favor ever-larger requirements on people to save and invest. The groups might also have a moral or ideological agenda, as in the case of temperance organizations (e.g., Mothers Against Drunk Driving) or personal health advocates (e.g., the Center for Science in the Public Interest). Importantly, these groups may not share the new paternalists' deference to the subjective preferences of targeted agents.

Third, once the initial policy is in place where no policy existed before, it often becomes politically cheaper than before to propose extensions to that policy. The logic of the political process often requires that a milder form of a policy be introduced and adopted first. There are at least three reasons for this: (a) The creation of an initial policy may involve incurring certain fixed costs, such as setting up an administrative agency to implement the policy. As a result, the added cost of extensions to that policy will be reduced.[35] For instance, the administrative cost of collecting a $1.25 tax per unit differs little from the cost of collecting a $1.00 tax per unit. (b) The attitudes of policymakers – voters, politicians, and judges – may change subsequent to the initial policy step, because even rational actors are subject to the "is-ought heuristic" which indicates that a rule or law that is in place provides evidence that this type of intervention is acceptable.[36] This effect is especially likely when it is costly to evaluate policy, and therefore voters and politicians rationally look to existing policies for signals of policy desirability. (c) Policymakers may be affected by the improved political position of an interest group that has a victory under its belt. It may, with good reason, appear to be more likely to win victories in the future.[37] Politicians want to hear "winning ideas" so they can claim to have made legislative accomplishments, which means interest groups with recent victories will be more likely to get a hearing. Together, these three factors (and possibly others) can make the legislative process susceptible to slipping down gradients.

The existence of a gradient does not, in itself, tell us which direction the sliding (if any) will go. In principle, we could imagine a slippery slope that leads toward less paternalism rather than more. The reason we predict asymmetrical movement toward greater paternalism is the coexistence of *other*, more specifically directional, slippery slope processes that we will discuss in the sections to follow. Gradients create fertile ground for those processes to operate.

b. How the New Paternalism Creates New Gradients

[34] See Lode (1999), 1513: "…[P]eople with power and influence also may stand to gain economically from taking steps down the slope. In addition, they may think that it is better from a moral point of view to take such steps."
[35] Volokh (2003), 1039-1051.
[36] Volokh (2003), 1081.
[37] Volokh (2003), 1122-23.

The new paternalist paradigm, as presented by its leading advocates, relies on discarding sharp distinctions in favor of gradients. Specifically, they reject standard distinctions between choice and coercion and between public and private action.

Cass Sunstein and Richard Thaler minimize the importance of the distinction between paternalism in the private and in the public sectors.[38] In explaining their concept of "libertarian paternalism," they say that the distinction between libertarian and non-libertarian paternalism "is not simple and rigid."[39] Moreover, they explicitly state that libertarian and non-libertarian paternalism lie on a continuum:

> The libertarian paternalist insists on preserving choice, whereas the non-libertarian paternalist is willing to foreclose choice. But in all cases, a real question is the cost of exercising choice, and here *there is a continuum rather than a sharp dichotomy*… [40]

Sunstein and Thaler thus present us with a gradient on which choice is characterized by low costs of escaping the prescribed course of action, while coercion corresponds to higher costs of escape. Who imposes the costs of escape and how they are imposed are regarded as unimportant questions; costs may be imposed by the state or private actors, with physical force and punishments or without.

In keeping with this framework, Sunstein and Thaler begin their analysis with a low-cost-of-exit point on the continuum: the seemingly innocuous question of where sugary desserts are placed in a cafeteria line.[41] If the fruit is placed before the cake and cookies, patrons are more likely to choose the former over the latter.[42] Sunstein and Thaler assume this to be in the best interests of all or most or some patrons, all things considered.[43] If the cafeteria owner chooses the placement in accordance with these best interests, he will, they claim, be acting *paternalistically.*[44] And since no coercion is involved, we see that paternalism need not preclude the patrons from making alternative choices. They can simply avoid the fruit and pick up the cake. No options are completely blocked, although the costs of exercising some of them are raised.[45] This is presented as a pure case "libertarian paternalism." (As Daniel Klein points out[46], the use of the term "libertarian"

---

[38] Sunstein & Thaler (2003), 1162: "…the same points that support welfare-promoting private paternalism apply to government as well."
[39] Sunstein & Thaler (2003), 1185.
[40] Sunstein & Thaler (2003), 1185, emphasis added.
[41] Thaler and Sunstein (2003), 175; Sunstein & Thaler (2003), 1164.
[42] We are not aware of evidence on this matter, but perhaps the illustration is just fanciful.
[43] It is unclear just what welfare claim is being made; that is, Sunstein and Thaler do not specify by what standard eating fruit is deemed superior to eating sugary desserts. To be consistent with their analytical perspective, the standard should not be based on health alone; taste should also matter.
[44] Presumably, if the best-interests placement were also the profit-maximizing placement, this would not be considered paternalism. In the case of consistency with profit maximization, the customers are getting what they explicitly want – a nudge toward the fruit.
[45] Specifically, a customer who puts fruit on his tray and then sees cake that he prefers will have to put back the fruit.
[46] Daniel B. Klein, *Statist Quo Bias*, 1 ECON JOURNAL WATCH 260 (2004).

in this context serves no analytical function, since it would be equally libertarian for private restaurants owners to eliminate desserts from the menu altogether; they have no obligation to provide desserts in any form to their customers.)

A similar example, with somewhat higher costs of exit, is the automatic enrollment of employees in retirement savings programs. This seems to increase savings for retirement, which Thaler and Sunstein consider a good thing, all things considered.[47] Employees, however, can opt out by signing the appropriate forms. So far there is no mention of coercion, by government or anyone else.[48] (Again, the word "libertarian" serves no analytical function here, since it would be equally libertarian for employers to require their employees to join a savings plan with no exit option, or even to pay lower wages and invest money on employees' behalf.)

Further along the gradient is the suggestion that a legal mandate on employers to adopt automatic enrollment may be consistent with "libertarian goals."

> For example, the law might authorize a situation in which employees have to opt into retirement plans, or it might require employers to provide automatic enrollment and allow employees to opt out. Both systems would respect the freedom of employees to choose, and either is libertarian in that sense.[49]

That the freedom of *employers* is restricted goes unmentioned. Nevertheless, the state has now entered the picture in a way that is ever so tentative – a relatively small intervention that is not explicitly advocated, but simply mentioned as a possibility.

The logic of the Thaler-Sunstein framework, however, implies compulsory automatic enrollment. This is because if employees recognized the benefits of automatic enrollment, employers who provided it would gain an advantage on the market. Under these conditions paternalism would not be necessary; profit maximization would be sufficient to achieve the increase in employee wellbeing.[50] Colin Camerer and coauthors, the other leading exponents of the new paternalism, recognize this and explicitly advocate legally mandated automatic enrollment.[51]

---

[47] But *see* Mario J. Rizzo, *Trust Us*, Forbes 30 (June 18, 2007), available at http://www.forbes.com/columnists/forbes/2007/0618/030.html?partner=whiteglove_google. Whether it is a good thing depends on the individual's personal circumstances. For example, for low and middle income earners fully participating in a 401K program (that is, from age 25) actually raise their lifetime tax payments and reduce their lifetime expenditures, because shifting of income to older age raises the portion of Social Security income subject to taxation, reduces the value of mortgage tax deductions, and raises the individual's marginal tax bracket in later years. *See* Jagadeesh Gokhale, Laurence J. Kotlikoff and Todd Neumann, *Does Participating in a 401K Raise Your Lifetime Taxes?*, National Bureau of Economic Research Working Paper No. 8341 (2001), available at http://www.nber.org/papers/w8341.

[48] Sunstein & Thaler (2003), 1172-1173.

[49] Sunstein & Thaler (2003), 1176.

[50] See note 42, above. In fact, recent data show that firms are adopting automatic enrollment voluntarily. This appears to be the result of recent legislation that has reduced the risk of law suits for channeling employees' wages into inappropriate investment vehicles**.** .Whether the percentage of firms voluntarily adopting automatic enrollment will ultimately satisfy new paternalists is yet to be seen.

[51] Camerer et al. (2003), 1252.

Still further along the continuum of coercion lie default or framing rules. These are cases in which the law must allocate a particular entitlement in the absence of any explicit agreement by the relevant parties. For example, if a labor contact says nothing about the conditions for termination, the law may presume "for cause" rather than "at will" termination. If this is the case, then the employer must buy the right to terminate at will from the employee, presumably in the form of higher wages.[52] Sunstein and Thaler claim that some amount of paternalism is "inevitable"[53] in cases like this, but that is not necessarily true.

Although it is true that there must be some default rule in cases like these, it does not follow that the rule must be chosen by the state with paternalist goals in mind. The law could merely accept the results of customary practice; what is recognized by the law need not have been created by it. The creation of a new default rule effectively *shifts transaction costs* (that is, the costs of negotiating agreements) from those who would depart from customary practice to those who would follow it. In the same vein, Sunstein and Thaler also suggest legal presumptions of guaranteed vacation time[54], specified on-the-job safety levels[55], and non-discrimination on the basis of age.[56]

How far such interventions to alter contractual defaults take us along the continuum of costs depends upon the ease of contracting out of the presumption. If there is no state-set price, the cost of contracting out may be relatively low; it is simply the transaction cost of inserting new terms in contracts. The next step along the continuum, however, is to impose *additional* transaction costs instead of merely shifting them. For example, Sunstein and Thaler offer the Age Discrimination in Employment Act (ADEA)[57] as another example of libertarian paternalism.[58] The ADEA allows employees to sign a "knowing and voluntary" waiver at retirement of age-discrimination protections. However, for the waiver to be valid, it must meet a series of requirements, including consultation with a lawyer, a twenty-one day waiting period, and a seven-day revocation period. Actually consulting a lawyer is costly for the employee, and the rest of the requirements are burdensome to the employer.[59]

Yet further along the continuum is legislation that creates "protections" for workers or consumers that can be waived, but *only under conditions set by the state*. Sunstein and Thaler offer the example of the Fair Labor Standards Act, which imposes a maximum number of hours per week, but allows the maximum to be waived.[60] In order to work

---

[52] However, the equilibrium wage under "for cause" would presumably be lower, meaning the net effect could be zero.
[53] Sunstein & Thaler (2003), 1173.
[54] Sunstein & Thaler (2003), 1175.
[55] Sunstein & Thaler (2003), 1175.
[56] Sunstein & Thaler (2003), 1176-1177.
[57] 29 USC sec. 626(f)(1) (2000).
[58] Sunstein & Thaler 2003, 1186-1187.
[59] Sunstein & Thaler (2003), 1187: "…the ADEA goes beyond the inevitable minimum level of paternalism by imposing those barriers which significantly raise the burdens of waiver."
[60] 29 USC sec. 207(f) (2000).

more than forty hours per week, workers must receive time-and-a-half pay for the extra hours. Employees cannot waive the maximum number of hours for any lower rate of pay. Here, the default rule is not merely a default, because it expressly prohibits certain exchanges. Similarly, Sunstein and Thaler point to the Model Employment Termination Act, which replaces "at will" with "for cause" termination. This right can be waived by agreement – but only if the employer agrees to provide a severance payment (equal to one month's salary for every year of employment) in the event of a not-for-cause termination. Of this policy, Sunstein and Thaler say that it is "less libertarian than it might be. But freedom of choice is nonetheless respected."[61] Yet freedom of choice is *not* fully respected; the employer and employee are prohibited from arriving at contracts with "at will" termination and no severance payment.[62]

Once we are in the realm of actually restricting the terms of voluntary agreements, the movement along the continuum is straightforward: either increase the cost associated with opting out, further restrict the terms of agreements, or both. Here, for instance, Sunstein and Thaler back mandatory cooling-off periods, during which consumers would be allowed to return purchased items (like cars) without penalty.[63] For this policy, they do not even mention the possibility of consumers waiving the requirement (even if doing so might earn a price discount).

At the far end of the continuum lies banning certain activities outright. Sunstein and Thaler embrace this conclusion:

> Almost all of the time, even the non-libertarian paternalist will allow choosers, at some cost, the reject the proposed course of action. Those who are required to wear motorcycle helmets can decide to risk the relevant penalty, and to pay it if need be.[64]

Notice that the same argument would place outright prohibition of alcohol, drugs, or anything else on the same spectrum. You are free to use any drug you want, says the argument, if you're willing to incur the cost of potential imprisonment. At this end of the continuum, we find, lies genuine hard paternalism. In Sunstein and Thaler's words:

> A libertarian paternalist who is especially enthusiastic about free choice would be inclined to make it relatively costless for people to obtain their preferred outcomes. (Call this a *libertarian* paternalist.) By contrast, a libertarian paternalist who is especially confident of his welfare judgments would be willing to impose real costs on workers and consumers who seek to do what, in the paternalist's view, would not be in their best interests. (Call this a libertarian *paternalist*.)[65]

---

[61] Sunstein & Thaler (2003), 1187.
[62] Employees might wish to make such agreements if the alternative is lower wages or unemployment. Ruling out some contract options means that some contracts won't be made at all.
[63] Sunstein & Thaler (2003), 1187-1188.
[64] Sunstein & Thaler (2003), 1189-90.
[65] Sunstein & Thaler (2003), 1185-86.

Movement along a paternalist continuum should come as no surprise when the two ends of the continuum depend on which word is italicized, as well as on the subjective confidence of the policymaker in his welfare judgments.
It bears emphasis that the sequence of steps we've outlined – from nudging (changing the order of cafeteria items) to pushing (imposing costs on those who deviate from the state's preferred terms of contract) to shoving (ruling out some terms entirely) to controlling (banning some activities altogether) – is not our creation. Sunstein and Thaler present the same proposals in approximately the same order, to demonstrate the existence of a continuum.

We have focused on Sunstein and Thaler's work because they are admirably explicit about their belief in a paternalist continuum. But the same pattern can be observed in Camerer and coauthors, who also structure their proposals in order from the seemingly innocuous to the fully intrusive. They summarize the progression like so:

> We focus on four types of policies: (1) default rules; (2) provision or re-framing of information; (3) cooling-off periods; and (4) limiting consumer choices. This list is ranked roughly in increasing order of departure from pure asymmetric paternalism – i.e., the increasing "heavy-handedness" of the policy.[66]

Again, we see that the leading new paternalists themselves believe that soft and hard paternalism can be connected by a series of small steps. Like Sunstein and Thaler, Camerer and coauthors present public and private, and coercive and non-coercive, paternalistic activities alongside each other with little or no recognition of when they are crossing the line from one to the other. In discussing "asymmetrically paternalistic" regulations that operate by requiring the provision of information, for instance, they offer state occupational licensing.[67] Unless they mean a form of licensing that merely requires the unlicensed to reveal that fact (a form of licensing for which we are hard pressed to find a single example), this classification is completely mistaken. Licensing requirements typically coerce both service providers and clients by preventing them from engaging in voluntary transactions – but the authors do not mention this.

It might be objected that the existence of a gradient from soft to hard paternalism is just a fact, and the new paternalists cannot be faulted for pointing it out. But the gradient in fact results from the conceptual framework that the new paternalists have adopted and urge the rest of us to adopt. The main problem with the framework, in our view, is that it defines paternalism (and libertarianism) in terms of costs of exit, *without any attention to who imposes the costs and how*. An alternative framework, one that is more consistent with the typical usage of words like coercion and choice, would focus on whether rights of person and property are abridged by a given policy. On this approach, a restaurateur's decision about dessert placement and a government's decision about whether to allow helmetless motorcycle riding simply would not be on the same continuum. The former is

[66] Camerer et al. (2003), 1224.
[67] Camerer, et al. (2003), 1237.

both private and non-coercive, the latter public and coercive. This is the sort of framework that the new paternalists encourage us to reject in favor of theirs.

c. How the New Paternalism Exploits Existing Gradients

In addition to creating new conceptual gradients, the new paternalism also exploits gradients that already exist as a result of theoretical or empirical vagueness.

i. Hyperbolic and Quasi-Hyperbolic Discounting

An important literature in behavioral economics holds that some, many, or even most people exhibit "excessive impatience" in important decisionmaking contexts. In their desire for short-run gratification, individuals may give too little weight to the possible longer-run consequences of their actions. This may result in insufficient savings for retirement, consumption of too much junk food, and so forth.

Traditional economic theory assumes that people's rate of trade-off or discounting between successive time periods is constant. For example, if an individual considers $100 to be received in two years equivalent to $90 received in one year, he should also consider $100 to be received in one year equivalent to $90 to be received now. This person is said to have a constant *discount factor* of 0.90. This is known as exponential discounting.[68] But real people appear to have inconsistent rates of discount: they exhibit higher rates of discount between time periods the closer those periods are to the present. For instance, an individual might consider $100 to be received in two years equivalent to $90 to be received in one year, but consider $100 to be received in one year to be equivalent to only $80 now. This is known as hyperbolic discounting.[69]

People with hyperbolic rates of discount exhibit *time inconsistency*: they will make decisions about future trade-offs, and then reverse those decisions later. For instance, consider a choice between $100 in two years and $85 in one year. The exponential discounter described above would choose the $100; and if offered the chance to reverse his decision after a year has expired (so that he is choosing between $100 in one year and $85 now), he would refuse. His choices are consistent. The hyperbolic discounter described above would also initially choose the $100; but if offered the chance to reverse his decision after a year has expired, he would do so. This is a result of his inconsistent rates of discount (0.90 between one and two years, but 0.80 between zero and one year).

In short, hyperbolic discounting means that people at first make long-term plans for saving or dieting but then, when the time comes to implement these plans, they succumb to the desire for short-run gratification. For the new paternalists, this type of behavior suggests an opening for paternalist intervention or correction. Examples include the

---

[68] It is called exponential the discount factor can be raised to a power equal to the number of time periods in question. For example, a one-period discount factor of 0.90 implies a two-period discount factor of (0.90)(0.90) = 0.81.

[69] The seminal article in this literature is Strotz (1955/56). Technically, we have just described quasi-hyperbolic discounting; the distinction between hyperbolic and quasi-hyperbolic will be clarified later.

previously mentioned proposal to automatically enroll people in savings plans[70], and to impose a sin tax (on unhealthy foods, cigarettes, and so forth) to provide additional incentive for impatient people to resist their temptations.[71]

New paternalists claim that they are evaluating the observed behavior of the individual in terms of his own normative standard.[72] This appears attractive until we realize that *the individual has no unambiguous standard* for the appropriate level of time discounting.

The analytical "opening" for paternalist policy is created by the existence of an internal inconsistency of choice. But although an inconsistency does create a quandary for traditional rational choice theory – which assumes that people have internally consistent preferences – it does not provide any grounds for *choosing* between the inconsistent preferences. The inconsistency of a hyperbolic discounter could be "fixed" by making him uniformly more patient (in the example above, always having an annual discount factor of 0.90), but it could also be "fixed" by making him uniformly less patient (always having an annual discount factor of 0.80).[73]

To craft new paternalist policies, it is necessary to decide the appropriate normative rate of time discounting. This matters because policy must specify the amount of money an individual is automatically signed up to save, the magnitude of a fat tax, etc. Which rate of discount is the correct one? Theory provides no answer, but the new paternalists have not hesitated to side with the more patient one. O'Donoghue and Rabin define "optimal behavior" as "that [which] maximizes long-run well-being," where long-run well-being is associated by the more patient rate of discount.[74] Gruber and Koszegi "take the agent's long-run preferences as those relevant for social welfare maximization."[75]

Abstractly, we might say that the normative rate is the one that arises out of a more considered deliberation of costs and benefits. But is the more patient rate really the result of a superior deliberative process? Suppose that the planning agent prefers a larger reward later (LL) over a smaller reward sooner (SS). Subsequently, the deciding agent switches and prefers SS over LL. For example, an individual plans to save more for retirement, and then *at the moment of deciding* between taking a vacation now and saving for the future arrives he chooses the former. All we know from a behavioral perspective is that the individual is exhibiting a time inconsistency, that is, he is changing his mind. Is

---

[70] Automatic savings plan enrollment is also justified by reference to status quo bias.
[71] See, especially, O'Donoghue & Rabin (2003a and 2003b).
[72] *See* ,for example, Sunstein & Thaler (2003), 1162: "[W]e emphasize the possibility that in some cases individuals make inferior decisions in terms of their own welfare – decisions that they would change if they had complete information, unlimited cognitive abilities, and no lack of self-control." Ibid., 1163: "The false assumption is that almost all people, almost all of the time, make choices that are in their best interests or at the very least be better, *by their own lights*, that the choices that would be made by third parties." Emphasis added. *See* also David I. Laibson, Andrea Repetto, Jeremy Tobacman, Robert E. Hall, William G. Gale and George A. Akerlof, *Self-Control and Savings for Retirement*, Brookings Papers on Economic Activity 91 (1998), 93: "People have a systematic tendency to err – as judged by their own standards – in the direction of instantaneous gratification."
[73] Whitman (2006), 5, 15, notes 17, 18.
[74] O'Donoghue and Rabin (2003), 5.
[75] Gruber and Koszegi (2001), 1287.

the rate of discount implicit in the planning more "considered" than the rate implicit in the ultimate decision? There are valid reasons to think the answer might be no. The deciding individual may actually have a better idea of the significance of the costs (or foregone benefits) at the moment of decision. The planning self may underestimate how re-invigorating a vacation or appealing a good dessert would be; for him, those sacrifices are purely notional. When the short-run benefits are closer at hand, he may have the benefit of superior local information.[76]

Thus, the normative standard inherent in any attempt to "help" agents with hyperbolic preferences is *inherently vague*. We do not know where "reasonable" impatience ends and "excessive" patience begins. There is no sharp dividing line between them. As we have argued above, such vagueness itself is sufficient to create a gradient.

But setting aside the theoretical vagueness, there is also empirical vagueness associated with using hyperbolic discounting to justify paternalist policymaking. This is true for two reasons. First, empirical estimates of actual rates of discount differ substantially.[77] It is conceivable that better future studies will isolate true rates of discount.[78] But we suspect that many of these confounding factors, such as changing utility and anticipatory utility, simply cannot be controlled. Furthermore, there are substantial differences across people in their degree of hyperbolic discounting.[79] Fernandez and Villaverde find, after controlling for the uncertainty that future rewards or penalties will actually obtain, only 13% of their sample are hyperbolic discounters, whereas other uncontrolled studies find that 40-60% are hyperbolic discounters.[80]

Second, and more importantly, there is empirical vagueness raised by the possibility that agents have multiple rates of discount, not just two. In the example above, the agent has only two different discount factors: 0.90 between any two successive years in the future, 0.80 between the present year and next year. Technically, this is known as *quasi-hyperbolic* discounting. But evidence indicates that people exhibit something closer to true hyperbolic discounting, in which there are multiple discount factors depending on

---

[76] "The information available to the acting –agent about the local consequences of a specific choice will often be better than the information available to the pre-agent [the planning agent]. When a dieter changes his mind and has tiramisu after promising not to, it might be because he is weak-willed, or it might be because he has only now realized how appealing the tiramisu is…" Daniel Read, *Which Side Are You On? The Ethics of Self-Command*, 27 JOURNAL OF ECONOMIC PSYCHOLOGY 681 (2006), 685.

[77] Shane Frederick, George Loewenstein, and Ted O'Donoghue, *Time Discounting and Time Preference: A Critical Review,* in George Loewenstein, Daniel Read and Roy Baumeister (eds), *Time and Decision*: *Economic and Psychological Perspectives on Intertemporal Choice* 13 (2003), 61.

[78] However, "there is no evidence of methodological progress in that the range of estimates does not seem to be shrinking with time." *See* Dilip Somin, George Ainslie, Shane Frederick, Xiuping Li, John Lynch, Page Moreau, Andrew Mitchell, Daniel Read, Alan Sawyer, Yaacov Trope, Klaus Wertenbroch and Gil Zauberman, *The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones?,* 16 MARKETING LETTERS 347 (2005), 354.

[79] *See* Somin et al. (2005), 354.

[80] *See* Jesus Fernandez-Villaverde and Arijit Mukherji, *Can We Really Observe Hyperbolic Discounting?*, unpublished manuscript at University of Pennsylvania (2006), 3.

how far a trade-off lies in the future.[81] For instance, the individual might have a discount factor of 0.95 for rewards to be received ten or eleven years in the future, a discount factor of 0.90 for rewards five or six years in the future; a discount factor of 0.80 for rewards two or three years in the future; and a discount factor of 0.70 for rewards now or a year from now. Behavioral economists have used quasi-hyperbolic instead of true hyperbolic discount functions not because of their correctness, but because of their analytical tractability[82]: they are easier to work with.

Quasi-hyperbolic discounting makes it deceptively simple to choose the "correct" rate of discount, since there appear to be only two options. If real people actually engage in hyperbolic discounting, this implies a gradient or continuum of discount rates over time. If we assume, notwithstanding our earlier objections, that the immediate discount rate is impulsive or ill-considered, which of the longer-term rates is normatively preferable? There is nothing in the logic of new paternalism or behavioral economics that can provide an answer. We are faced with a continuum of normative possibilities.

These arguments impel us to the conclusion that among the discount rates revealed in choice or planning behavior, none has a clear claim to normative superiority. Thus, the new paternalist is in a conceptual fog because his underlying standard of evaluation is unspecified. The notion of "excessive impatience" is both theoretically and empirically vague, and that means we have a gradient of possibilities. There is no clear line to resist the gradual creep of higher savings requirements, higher fat taxes, and the like.

ii. Framing and Context-Dependence

For a variety of decisions, people are subject to what behavioral economists call context-dependence. This means that how they choose among two or more options depends on seemingly irrelevant aspects of how the situation is described. For example, medical patients are more likely to assent to a treatment with a 90% survival rate than one with a 10% death rate, even though these are the same thing.[83] In this case, people seem to favor a "positive" over a "negative" framing. People also seem to prefer options framed as the existing or a baseline position; this may be called status-quo bias.[84] Another example of the power of framing is the persistent difference between willingness-to-pay (WTP) and willingness-to-accept (WTA), meaning that people will demand more money

[81] Richard Thaler finds three effective annual discount rates ranging from 345 percent over a one-month horizon to 120 percent over a one-year horizon to only 19 percent over a ten-year horizon (Richard H. Thaler, *Some Empirical Evidence on Dynamic Inconsistency*, 8 ECONOMIC LETTERS 201 (1981). The discount rate (r) is the percentage reduction per unit of delay (say, per year) in the value-when-received of a reward or penalty. It is related to the discount factor (δ) in the following way: $r = (1- \delta)/\delta$. For further references, see George Ainslie, *Précis of Breakdown of Will*, 28 BEHAVIORAL AND BRAIN SCIENCES 635 (2005), 636-7.

[82] George-Marios Angeletos et al., *The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Estimation*, 15(3) J. ECON. PERSP. 47 (2001), 50.

[83] Sunstein & Thaler (2003), 1161, 1179.

[84] Russell Korobkin, *The Status Quo Bias and Contract Default Rules*, 83 CORNELL L. REV. 608 (1998).

to part with an item than they will pay to acquire it, even when the item's value is a trivial portion of their wealth or income.[85]

The phenomenon of context-dependence underlies various new paternalist proposals. All of Sunstein and Thaler's proposals for new contractual defaults, for example, rely on the difference between WTP and WTA.[86] Although such defaults leave all contractual options open (at least for the most modest proposals), employees may be less willing to part with a given term (such as guaranteed paid vacation) than to bargain for its inclusion. If there were no difference between WTP and WTA, and if transaction costs were zero, then the realized terms of contract would be the same regardless of the default.

The problem with context-dependence is similar to that of hyperbolic discounting: the new paternalist argument relies on an internal inconsistency to justify intervention, but there is no theoretical basis for choosing which behavior represents the individual's "true" best interest as he sees it. Which better represents a person's real preferences: what he is willing to pay for something or what he is willing to accept to part with it? There is no theoretically correct answer to this question, as Sunstein and Thaler admit: "If the arrangement of alternatives has a significant effect on the selections the customers make, then their true 'preferences' do not formally exist."[87]

In the absence of a true underlying preference as the correct standard, what standard should be used? Sunstein and Thaler decline to answer that question: "We are not attempting to say anything controversial about welfare, or to take sides in reasonable disputes about how to understand that term."[88]

In short, there is no standard provided by behavioral economic theory. The answer to the "what standard" question will depend on policymakers' own particular notions of welfare and well-being, as well as the weight they attach to autonomy. Notably, behavioral economics does not necessarily place *any* weight on autonomy, despite Sunstein and Thaler's obeisance to the value of individual choice. Policymakers who adopt the new paternalists' approach need not share their belief in choice. The new paternalist paradigm places them on a gradient from policies that only mildly restrict choice to policies that restrict or abolish it.

While some paternalist policies are meant to correct context-dependence, others use context-dependence as a device to correct different biases. Christine Jolls and Cass Sunstein, citing research that over-optimistic consumers underestimate the risks associated with dangerous products[89], propose to correct the optimism bias by exposing consumers to stories or narratives about people who have used such products:

---

[85] Sunstein & Thaler (2003), 1177.
[86] Sunstein & Thaler (2003), 1181: "A default rule might create a 'pure' endowment effect. It is well known that people tend to value goods more highly if those goods have been initially allocated to them than if those goods have been initially allocated elsewhere. And it is well known that, in many cases, the default rule will create an initial endowment effect."
[87] Sunstein & Thaler (2003), 1164.
[88] Sunstein & Thaler (2003), 1163, note 17.
[89] Jolls & Sunstein (2006), 204-205.

> Specifically, the law could require firms – on pain of administrative penalties or tort liability – to provide a truthful account of consequences that resulted from a particular harm-producing use of the product, rather than simply providing a generalized warning or statement that fails to harness availability.[90]

The tendency of people to take narratives more seriously than statistical information is another form of context-dependence, known as availability bias. For the narratives to be effective, they have to be sufficiently frightening or visceral. And therein lies the problem: there is no objective line between "not frightening enough" and "too frightening."

Jolls and Sunstein admit that excessively frightening narratives could be counterproductive, inducing too little risk taking, and their response is telling: "Of course there are line-drawing problems here, but the basic point is straightforward."[91] Line-drawing problems are, of course, the telltale sign of a gradient. As an empirical matter, there is simply no way to know whether customers who engage in a risky activity are doing so rationally – with a full understanding of the risks – or have simply not been exposed to a sufficiently scary narrative. The gradient goes from missing narrative to mildly compelling narrative to worst-case-scenario narrative. Courts of law asked to adjudicate "insufficient narratives" claims under Jolls and Sunstein's proposed law could easily, if guided by precent-based decisionmaking, slide down the gradient.

The gradient could even cross the threshold between truth and falsehood. Although the policy description above specifies a "truthful account," there is no particular reason, in theory, to think a truthful account provides the appropriate visceral response to approximate a rational assessment of risk. Jolls and Sunstein themselves state that truthful information can, in cases, be harmful: "In the face of health risks, for example, some presentations of accurate information might actually be counterproductive, because people might attempt to control their fear by refusing to think about the risk at all."[92] If too much information is a bad thing, then policies that require withholding of information could be justified on the same grounds as policies that require providing narratives. (Imagine, for instance, the case of an HMO that presented its patients with scary – but truthful – information about the risks of certain costly treatments.) And if withholding information can be the correct choice, it might also be appropriate to lie – if such lies do a better job of pushing people toward what policymakers think are their best interests.

What Jolls and Sunstein propose, then, is a movement from a bright-line liability rule to provide truthful statistical information to liability based on a gradient, with no objective means, in theory or evidence, of saying what is correct.

---

[90] Jolls & Sunstein (2006), 212.
[91] Jolls & Sunstein (2006), 214.
[92] Sunstein & Thaler (2003), 1183.

5. SLIPPERY SLOPES WITH RATIONAL POLICYMAKERS

In this section we will analyze "rational" slippery slope processes, by which we mean those in which the choices of experts, policymakers, and rent-seekers are rational in the standard economic sense of the term. Only the choices of the targeted agents are subject to cognitive or behavioral biases. We adopt this approach to show that new paternalist policies have an expansive tendency *even if* policymakers are somehow immune to the cognitive and behavioral biases that new paternalists ascribe to most people.

In some cases, the slope process may be driven by decision-making heuristics, and some of these heuristics might be characterized as less than fully rational. However, in this section we only invoke those heuristics that can reasonably be understood as rational responses to scarcity of information and time. We will discuss three major categories of rational slippery slopes: altered incentive slopes; authority, simplicity, and distortion slopes; and expanding justification slopes .

Later, we will show that if cognitive or behavioral biases characterize other actors in the system – particularly policymakers – the slippery slope potential of new paternalism is intensified. Irrationalities grease the slope.

a. Altered Incentives Slopes

Sometimes acceptance of an initial argument or policy can induce unintended or unexpected changes in the behavior of target agents, thereby creating an incentive for policymakers (either the same policymakers or later ones) to enact further policies to control or correct the new behavior.

We focus on how new paternalist policies have the potential to create unintended consequences that encourage further intervention. This may occur through the interaction of the biases of the targets, the crowding out of targets' self-regulatory behavior, and the substitution between targets' personal inputs, all of which can impede attainment of the paternalists' goals.

i. Bias Interactions

Studies of cognitive and behavioral biases typically focus on one bias at a time. The issue of the interaction of individual biases rarely arises. Yet interactions are crucial for evaluating the impact of corrective policies, because in the uncorrected state, biases may partially offset each other. If one bias is corrected while another is ignored, the latter bias may create new problems or worsen existing ones. As Gregory Besharov has pointed out, this means that paternalist interventions can be welfare-reducing if bias interactions

are not properly taken into account.[93] Our concern, however, is the resulting slippery-slope potential: if paternalist interventions create or exacerbate other behavioral problems, policymakers will have incentives to engage in further interventions.

One simple example, provided by Besharov, is that people's overestimation of their future consumption needs can, at least partially, offset their tendency to undersave because of hyperbolic discounting.[94] If a new paternalist policy succeeds in correcting the former (forecasting) bias, it will exacerbate the (latter) motivational bias.

Another example is provided by fat taxes and cigarette taxes. An implicit assumption in these policy proposals is that the higher prices will be experienced by the current self – the one that is subject to the temptations of overeating and smoking. If, however, the targeted agents have access to credit, then they can "offload" the increased financial burden to their future selves.[95] To put it another way, there can be an interaction between present-bias in eating/smoking decisions and present-bias in savings decisions. An attempt to correct the former can therefore exacerbate the latter, which will give policymakers added incentive to regulate savings decisions.

If cognitive and behavioral biases interact in a simple manner – as in the simple paired interactions we have just discussed – then there may be a natural stopping point to the slippery slope. Corrective legislation in area A leads to a problem in area B, and thus to corrective legislation in area B, and the process ends. But if, as we suspect, cognitive and behavioral biases interact through a complex web of effects, then the process need not have a stopping point. Each corrective intervention leads to problems that potentially justify yet more interventions.

ii. Crowding Out of Self-Regulation

Bias interaction, as described above, can create slopes wherein intervention in one area created incentives for policymakers to intervene in other areas. Here, we consider a process by which intervention in one area creates an incentive for further intervention in the same area.

The effect occurs because the initial policy turns out to be ineffective or counterproductive for some reason, and therefore the same motivation that justified the initial interventions can justify further interventions. As Eugene Volokh has noted, policymakers will often cite the need to enforce an existing policy as reason for enacting new policies.[96] More broadly, enactment of a policy generally involves a commitment to achieving certain goals; as long as those goals remain unachieved, policymakers have the

---

[93] Gregory Besharov, *Second-Best Considerations in Correcting Cognitive Biases*, 71 SOUTHERN ECONOMIC JOURNAL 12 (2004).
[94] Besharov (2004), 12-13, citing Matthew Rabin, *Comment*, in BEHAVIORAL DIMENSIONS IN RETIREMENT ECONOMICS, Henry Aaron, ed., Brookings Institution Press and Russell Sage Foundation (1999).
[95] Whitman (2006), 11, 12.
[96] Volokh, 1051ff.

incentive to intervene further. In the case of new paternalist legislation, we think this effect most likely to arise from the crowding out of self-regulation.

People have various means of controlling their own cognitive and behavioral problems. Examples include making resolutions and commitments[97], using mental budgets to limit the extent of certain activities[98], imposing self-reward and self-punishment schemes[99], and creating self-imposed constraints[100] by (for instance) buying food in smaller quantities[101] and voluntarily accepting deadlines.[102] They may also deliberately expose themselves to external social controls, including the criticism of family and friends, as well as "[s]ocial partners, groups and organizations…institut[ing] incentives, sanctions and rules that are designed to help individuals overcome temptations."[103]

Experimental studies by Ayelet Fishbach and Yaacov Trope[104] have shown that the imposition of some forms of external control can cause individuals to engage in less "counteractive self-control." For example, when students were asked to take a boring "diagnostic test of their reading skills," they exercised counteractive self-control when there was no external control or pressure. Specifically, they increased their *ex ante* evaluation of the test's value relative to their evaluation of the same test characterized as "interesting." Thus the short-run costs of the test (boredom) were counteracted by viewing the test in a more positive light. However, when there was social monitoring in the form of having the experimenter present when subjects were deciding to take the test, the subjects did not view the boring test as more valuable. Nevertheless, they still

---

[97] Ainslie (2001), 90-95; Jon Elster, ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY (1984), 256-60.

[98] Chip Heath & Jack B. Soll, *Mental Budgeting and Consumer Decisions*, 23 JOURNAL OF CONSUMER RESEARCH 40 (1996); Richard H. Thaler, *Mental Accounting and Consumer Choice*, 4 MARKETING SCIENCE 199 (1985).

[99] David Glen Mick, *Giving Gifts to Ourselves: A Greimassian Analysis Leading to Testable Propositions*, in MARKETING AND SEMIOTICS: SELECTED PAPERS FROM THE COPENHAGEN SYMPOSIUM, ed. Hanne Hartvig Larsen, David Glen Mick, and Christian Alsted, Copenhagen: Hendelshojslolens, 142 (1991); David Glen Mick, *Self-Gifts*, in GIFT GIVING: A RESEARCH ANTHOLOGY, ed. Cele Otnes and Richard F. Beltramini, Bowling Green, OH: Bowling Green State University 99 (1996); David Glen Mick & DeMoss, Michelle, *Self-Gifts: Phenomenological Insights from Four Contexts*, 17 JOURNAL OF CONSUMER RESEARCH 322 (1990); Albert Bandura & Bernard Perloff, *Relative Efficacy of Self-monitored and Externally Imposed Reinforcement Systems*, 7 JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY 111 (1967); Albert Bandura & Dale H. Schunk, *Cultivating Competence, Self-Efficacy, and Intrinsic Interest through Proximal Self-Motivation*, 41 JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY 586 (1981).

[100] Strotz (1955-6), 173; Thomas Schelling, *Egonomics, or the Art of Self-Management*, 68 AMERICAN ECONOMIC REVIEW 290 (1978); Thomas Schelling, *Self-Command in Practice, in Policy, and in a Theory of Rational Choice*, 74 AEA PAPERS AND PROCEEDINGS 1 (1984).

[101] Klaus Wertenbroch, *Consumption Self-Control by Rationing Purchase Quantities of Virtues and Vice*, 17 MARKETING SCIENCE 317 (1998).

[102] Dan Ariely & Klaus Wertenbroch, *Procrastination, Deadlines and Performance: Self-Control by Precommitment*, 13 PSYCHOLOGICAL SCIENCE 219 (2002).

[103] Ayelet Fishbach & Yaacov Trope, *The Substitutability of External Control and Self-Control* 41 JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY 256 (2005).

[104] Fishbach & Trope (2005).

decided to take the test as a result of the social pressure from experimenter's monitoring. Thus, self-control and social pressure behaved as substitutes.[105]

In another of Fishbach and Trope's studies[106], students were offered a "highly valuable" diagnostic test of their nighttime cognitive abilities. The test would be administered either at 9pm or at the more inconvenient time of 1am. Half the students would be given a payment of $20 to take the test while the other half would not receive any payment. The results were similar to those in the previously-mentioned study: external control diminished the extent of counteractive self-control. Specifically, the unpaid participants rated the test's importance more positively than did paid participants, and they were also willing to impose a greater punishment on themselves (in the form of a cancellation fee) for failing to take the test.

If individuals perceive external control and self-control as substitutes, this has important implications for paternalist policymaking. Suppose a policymaker decides to place a tax on a "bad" activity or a subsidy on a "good" activity. Insofar as the new policy is perceived as a form of external control, the evidence above suggests that targeted agents will respond by decreasing their level of self-control. This means that if such results are difficult to anticipate, as they will be for rationally ignorant policymakers, the initial policy will be later found insufficient. Then arguments will be made for increasing the subsidy or tax and expanding the degree of paternalistic intervention.[107] The changing behavior of the target agents generates a new round of incentives for policymakers to intervene.

There is a close analogy here to the economic literature on public goods and externalities. Economic theory indicates that an increase in the amount of state funding for a public good (say, maintenance of a public park) can lead to an offsetting decrease in the amount of private funding for the same good.[108] Similarly, James Buchanan has argued that a tax on a negative externality (his example is the noise nuisance created by someone's barking dogs) can decrease the decree to which the nuisance-creator takes into account costs to others, like his neighbors, when deciding how much nuisance to create.[109] In a sense, the state-imposed tax diminishes the self-imposed tax. The same point applies in the context of paternalist policies designed to correct "internalities" created by conflict between

---

[105] Fishbach & Trope (2005), 260-261. Similarly, when students were asked to evaluate studying – an activity with short-run costs and long-run benefits – counteractive self-control and external control in the form of parental expectations were substitutes in overcoming the temptations of interfering activities like watching TV. Fishbach & Trope (2005), 261-263.

[106] Fishbach & Trope (2003), 263-266.

[107] Economists will no doubt prefer models of a social optimum in which there is an equilibrium level of a subsidy (tax) corresponding to an equilibrium level of self-control. If such a model were applicable, then if the policymaker knew the socially optimal subsidy he could simply impose it and the socially optimal level of self-control would simply be generated. The reality of policy is, however, far more messy.

[108] See Burton Abrams & Mark Schmitz, *The Crowding Out Effect of Government Transfers on Private Charitable Contributions*, 33 PUBLIC CHOICE 29 (1978); B. Douglas Bernheim, *On the Voluntary and Involuntary Provision of Public Goods*, 76 AMERICAN ECONOMIC REVIEW 789 (1986); Theodore Bergstrom, Lawrence Blume, and Hal Varian, *On the Private Provision of Public Goods*, 29 JOURNAL OF PUBLIC ECONOMICS 25 (1986).

[109] James Buchanan, COST AND CHOICE (1969).

one's present and future selves: we can expect that individuals will reduce the extent of their self-regulation.[110]

iii. Substitutability of Personal Inputs

In the previous section we saw that external and internal self-control can be substitute inputs in the attainment of one's goals. Here, we note that different forms of personal activity can also be substitute inputs. For instance, the goal of reducing one's weight can be served by two activities: eating healthily and exercising regularly. Following Fishbach, Dhar, and Zhang[111], we will refer to the larger goal as the "superordinate goal" and the inputs as "subgoals." To the extent the individual views the subgoals as substitutes, greater attainment of one subgoal will lead to a reduction in the other subgoal.

Now suppose that some form of external control leads to greater attainment of a subgoal. For instance, a fat tax helps the individual to improve his diet, and this might seem to be an improvement (from a paternalist perspective). However, he might also reduce his effort on other subgoals, like exercise. The overall effect on the individual's weight could be negligible or even positive. As a result, policymakers have reason to intervene further – by increasing the fat tax, or subsidizing gym memberships, or perhaps even resorting to food bans and exercise mandates.

The above reasoning follows even in a straightforward rational-choice framework, given personal preferences that regard healthy eating and frequent exercise as substitutes in the maintenance of weight. But if targeted agents are susceptible to a myopic focus on subgoals, rather than focusing on the superordinate goal, this makes the substitution effect even more likely.[112] When a person is focused on the *subgoal* of eating more healthily, he is likely to interpret the effect of the fat tax as progress toward the primary goal of good health. Then, in a world of multiple goals and a desire to maximize overall utility, he may shift away from exercise to the pursuit of other, unrelated goals – such as the relatively neglected goal of watching television. [113]

If the individuals affected by the tax are relatively myopic, insofar as their focus is mainly on the subgoal of reducing the junk or fat content of their diet, this research suggests that they will cut back on other health-producing activities such as working out.

---

[110] Whitman (2006), 12-13.

[111] Ayelet Fishbach, Ravi Dhar, & Ying Zhang, *Subgoals as Substitutes or Complements: The Role of Goal Accessibility*, 91 JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY 232 (2006).

[112] "The degree to which individuals interpret subgoal attainment in terms of progress or commitment depends on their attention to the relatively concrete subgoal in comparison to the corresponding, relatively abstract superordinate goal…. In the course of pursuing multiple goals (e.g., weight loss and food enjoyment), partial fulfillment of a focal goal suggests to the individual that other objectives are somewhat neglected, further motivating disengagement from that goal." Fishbach, Dhar, & Zhang (2006), 233.

[113] "On the other hand, when the focus is on the superordinate goal, the same level of successful attainment highlights commitment to the overall goal….For example, when the goal to have an attractive figure is highly accessible, an initial success at losing weight strengthens the commitment to this goal as well as related activities toward that end (e.g., working out)." Fishbach, Dhar, & Zhang (2006), 233.

Whether a tax on the fat content of food actually promotes greater health will be affected not only by the impact on fat intake but also by the impact on complementary activities. It is not a foregone conclusion that the net effect will be positive. Indeed, if individuals are being "compelled" to reduce fat intake by a tax when they have no operative interest in doing so, then it is likely that their main focus will be on the subgoal, meaning the substitution effect is more likely.

More generally, when there are multiple inputs to an overarching goal, paternalistic policies that increase input to that goal can decrease the other input, thereby making the policy less effective or even counterproductive. As a result, policymakers may be tempted to engage in further interventions.

b. Authority, Simplification, and Distortion Slopes

Substantial deference to authority is inherent in the application of new paternalist ideas to public policy. This is because the complexities, vagueness, and indeterminism of their analysis (previously discussed) raise the costs of decisionmaking on the part of voters, politicians and bureaucrats. The locus of effective decisionmaking will then quite reasonably shift to experts ("authorities") or to simplifiers of technical ideas who may have agendas of their own. As Eugene Volokh puts it, "The more complicated a question seems, the more likely it is that voters will assume that they can't figure out for themselves and should therefore defer to the expert judgment of authoritative institutions…"[114]

There will thus be a tendency for policy to slide away from the values of the targeted agents themselves toward those of outsiders regarded as authorities. This happens in at least two ways. First, experts simplify their own theories to make them applicable in an policy context. Second, people seeking to advance their own interests will further simplify the theory and distort the facts to suit their purposes.

i. Simplification of Theory by Experts

Although it may seem as if the shift of effective decisionmaking to experts is just the right thing in difficult cases, this is not always true. It is especially unlikely to be true in the case of new paternalist policies. This is because, as we have argued earlier, the underlying standards and information needed to apply those standards and implement policy are *fundamentally* vague and indeterminate. The experts themselves have, at best, only a tenuous grip on the values of the targeted agents, which limits the direct applicability of their paternalistic theories to policy. Thus, there will be a tendency for the experts to reify their own values, and to simplify their own theories, in order to make more definite policy recommendations. We offer two examples of how this can occur, or has already occurred, in the new paternalist literature.

Resolving internal conflict in favor of the expert's preference. Consider the previously mentioned assumption, on the part of analysts trying to calculate optimal sin taxes, that the appropriate normative rate of time discounting is the longer-run rate (that

[114] Volokh (2003), 1082.

is, the one that applies when all costs and benefits lie in the future).[115] This assumption is offered without argument, but the choice is probably not arbitrary. It reflects, no doubt, certain intellectual middle-class values – not coincidentally the values of many experts. A similar presumption underlies the assumption that, between decisionmaking in a "hot" state (of fear, anxiety, arousal, etc.) and decisionmaking in a "cool" state (calm and sober), the latter must better represent an agent's "true" preferences.[116]

The underlying values of the new paternalists are also revealed in their more "casual" statements of their positions. Thaler and Sunstein, along with many others, think that Americans are too fat. Too fat by what standard? As Thaler and Sunstein recognize, consistent new paternalism requires that the standard be the *overall* preferences of the targets themselves. Obviously, these preferences involve a certain balancing of health risks, attractiveness, enjoyable food, avoidance of unpleasant physical exertion, and so forth – all in the context, of course, of ultimate certain death. Yet their evidence for the proposition that Americans are too fat is simply that Americans' weight has increased despite the health consequences: "Given the adverse effects obesity has on health, it is hard to claim that Americans are eating optimal diets."[117] This is a complete non-sequitur. There are costs of being overweight; no one is denying these.[118] But there are benefits of indulgence as well. In a later paper Sunstein and Thaler seem to recognize the subjectivity of the cost-benefit balance:

> Of course, rational people care about the taste of food, not simply about health, and we do not claim that everyone who is overweight is necessarily failing to act rationally. It is the strong claim that all or almost all Americans are choosing their diet *optimally* that we reject as untenable.[119]

What does "optimal" look like? Even if everyone acted in a manner fully consistent with their "true" preferences, there would surely be some people who qualified as overweight or obese. People would still die of weight-related illnesses, so the mere fact of such illnesses does not constitute evidence of suboptimal diets.

When we lack good information about the target agents' underlying, but unrevealed, preferences, the notion of optimality takes on an ethical or first-person dimension. If the evidence reveals a conflict between competing preference orderings within the individual, neither theory nor evidence provides a basis for favoring one preference ordering over another. Yet the new paternalists do not hesitate, when making policy recommendations, to choose among the competing preferences.

---

[115] Supra notes 69, 70.
[116] George Loewenstein (2000), *Emotions in Economic Theory and Economic Behavior*, 90 AMERICAN ECON. REV. 426; Sunstein & Thaler (2003), 1188: "The essential rationale [for cooling-off periods] is that under the heat of the moment, consumers might make ill-considered or improvident decisions."
[117] Thaler & Sunstein 2003, 1167.
[118] There is, however, some doubt as to how overweight one must be for the ill health effects to occur. See infra notes 128-131 and accompanying text.
[119] Sunstein & Thaler (2003), 1168.

The new paternalists claim to have found policy interventions that will make targeted agents better off according to the *target agents*' own preferences. What they have in fact found is evidence of internal conflict in the target agents' preferences, and then they have resolved the conflict in favor of the *experts*' preferences. The error in reasoning is subtle enough that the experts themselves have simplified the argument substantially – either because they do not fully understand the argument themselves, or because they do understand the argument but have simplified it for mass consumption.

What creates the slippery-slope potential here is the veneer of scientific objectivity. It is the simplified argument, not the original and more sophisticated one, that becomes reified in policy. Yet the simplified form of the argument can justify far more than the initial intervention, especially if the experts are appointed to agencies and commissions tasked with implementing it. If simple observations – that people weigh more than they used to, that they don't save as much as we think they should – are taken as *ipso facto* evidence of suboptimal choices, then further intervention will surely follow.

Eliding important qualifications of new paternalist arguments. Consider the rule for asymmetric paternalism proposed by Camerer and coauthors.[120] If some fraction of the public p is irrational, irrational people will receive a per capita benefit of B, and rational people will suffer a per capita cost of C, then the policy is justified if

$$pB - (1 - p)C > 0$$

(We have simplified their model slightly to exclude implementation costs and profits to firms.) The rule corresponds to a set of justifications: if the benefit (B) to irrational people is large enough, *and* the cost to rational people (C) is small enough, *and* the ratio of irrational to rational people is high enough, then the policy is justified. Notice, however, the shorthand explanation the authors offer:

> Such policies are appealing because, even possessing little information about the frequency of consumer errors, as long as we think p is positive – as long as we can get even the truest believer in consumer rationality to concede that some agents, some of the time, exhibit bounded rationality – we can conclude with some confidence that the policy is on net beneficial.[121]

This explanation is wrong. Given their own criterion, it is not sufficient for the fraction of irrational people to be positive; it must also be true that the ratio of benefits (B) to costs (C) is greater than the ratio of rationals to irrationals. To the authors' credit, in the very next paragraph they clarify the principle.[122] We have less confidence, however, that policymakers internalizing the asymmetric paternalism criterion will be so careful. Instead, they could easily draw out a simpler principle: *a policy is beneficial as long as there exist some irrational people*. This is a case of a broader justification substituting for a narrower one (or alternatively, a stripping away of important qualifications). Even

[120] Camerer et al (2003), 1219.
[121] Camerer et al. (2003), 1219.
[122] Camerer et al. (2003), 1219-1220.

if the initial policymakers do not employ this broader principle, subsequent policymakers might well infer it.

To compound this problem, however, the narrower, initial criterion is itself a significant simplification, in that it imagines that the population can be cleanly divided into rational and irrational groups. In reality, irrationality presumably exists on a spectrum, with people being subject to varying degrees of irrationality. Thus the proportion of people significantly benefitting from a policy may be smaller than the size of the "irrational" category would suggest. The new paternalists already felt the need to elide such complications in order to explain their approach to *intellectuals*. We should not be surprised to observe much greater simplifications to occur when policies are being sold to voters and politicians and implemented by bureaucrats.

ii. Distortion of Facts by Rent-Seekers

We have argued that, even if expert opinions lead directly to policy outcomes, there is the potential for a slippery slope. But in the real world of public policy, expert recommendations do not translate directly into policy. Instead, their opinions are filtered, amplified, and simplified by non-experts with agendas of their own. We refer to the latter group as rent-seekers. We offer two examples.

Exaggeration of risks from second-hand smoke. The rent-seekers' distortions are especially evident in the controversy over the effects of second-hand smoke and the related policy issue of banning indoor smoking in workplaces and restaurants or even in outdoor spaces. There has been widespread simplification and distortion of the scientific evidence by anti-tobacco groups.

We should note that although exposure to second-hand smoke ("environmental tobacco smoke" or ETS) is not a strictly paternalist context, inasmuch as second-hand smoke can potentially harm bystanders, paternalist arguments have played an important supporting role. Most importantly, note that many actual and proposed anti-smoking regulations limit the ability of individuals who may not be bothered by smoke to expose themselves *voluntarily* to second-hand smoke as customers or employees of restaurants and bars. Furthermore, by creating a hostile environment for smokers, the ETS argument easily slides into the paternalistic. Thus, even some ETS arguments must be regarded as partially paternalistic in either intention or merely in effect.

To understand the nature of the simplification-distortion process, consider the following statements. The first is by a Washington, D.C. anti-smoking advocacy group, Action on Smoking and Health (ASH):

> Even for people without such respiratory conditions [asthma, chronic bronchitis, chronic sinusitis, etc.] breathing drifting tobacco smoke for even brief periods can be deadly. For example, the Centers for Disease Control [CDC] has warned that breathing drifting tobacco smoke for *as little as 30 minutes* (less than the time one

> might be exposed outdoors on a beach, sitting on a park bench, listening to a concert in a park, etc.) can raise a nonsmoker's risk of suffering a fatal heart attack to that of a smoker.[123]

The second is by another prominent anti-smoking group, SmokeFreeOhio:

> After *twenty minutes of exposure* to secondhand smoke, a nonsmoker's blood platelets become as sticky as a smoker's, reducing the ability of the heart to pump and putting a nonsmoker at an elevated risk of heart attack.[124]

The first statement purports to be based on a CDC study which, in fact, does not support the claim[125], while the second statement is a partial truth that does not support the conclusion about heart-attack risk.[126] The kernel of truth in both statements is that exposure to ETS *over a long period of time* would produce atherosclerosis and thus an elevated risk of heart attack; but the underlying scientific literature does not provide evidence for the same effects resulting from ETS exposure of twenty or thirty minutes. Thus, a long-term risk, without the temporal qualification, is simplified into an immediate risk.[127]

---

[123] Action on Smoking and Health (ASH), *Reasons for Banning Smoking in Certain Public Outdoor Areas*, available at http://ash.org/outdoors.pdf, accessed April 1,2008, emphasis added. Cited in Michael Siegel, *Is the Tobacco Control Movement Misrepresenting the Acute Cardiovascular Health Effects on Secondhand Smoke Exposure? An Analysis of the Scientific Evidence and Commentary on the Implications for Tobacco Control and Public Health Practice*, 4 EPIDEMIOLOGIC PERSPECTIVES AND INNOVATIONS 1 (2007), 2, available at http://www.epi-perspectives.com/content/pdf/1742-5573-4-12.pdf, accessed April 1, 2008.

[124] SmokeFreeOhio, *The Dangers of Secondhand Smoke*. Available at http://www.smokefreeohio.org, emphasis added. Cited in Siegel (2007), 3.

[125] See Siegel (2007), 5-8. The CDC study (actually a study by two authors at the CDC) did not consist of original research but was an analysis of the literature. It concluded that "even 30 minutes of exposure to a typical dose of secondhand smoke induces changes in arterial endothelial function [an aspect of coronary circulation] in exposed non-smokers of a magnitude similar to those measured in active smokers." See T.F. Pechacek and S. Babb, *Commentary: How Acute and Reversible are the Cardiovascular Risks of Secondhand Smoke?* 328 BRITISH MEDICAL JOURNAL 980 (2004), 981. This statement quoted was, in turn, based on a study conducted by R. Otsuka et al., *Acute Effects of Passive Smoking on the Coronary Circulation in Healthy Young Adults*, 286 JAMA 436 (2001). How does the actual claim of these articles relate to the assertion of ASH quoted in the text above? "What this [actual finding] means is that acute exposure to secondhand smoke can result in endothelial dysfunction in nonsmokers that if prolonged and repeated over a long time, could eventually result in atherosclerosis and heart disease. This study provides a potential mechanism for the observed increase in heart disease risk among involuntary smokers. It provides biologic plausibility for a causal relationship between chronic exposure to secondhand smoke and heart disease. But it does not suggest that an otherwise healthy nonsmoker could suffer a heart attack as a result of a 30 minute exposure to secondhand smoke, and *it certainly does not mean that a nonsmoker's risk of a heart attack approaches that of a smoker after 30 minutes of exposure to secondhand smoke*." Siegel (2007), 6-7, emphasis added.

[126] For a critical analysis, see Siegel (2007), 9-13, especially p. 10.

[127] This overall picture is, in fact, quite clear. See Siegel (2007), 24: "While there is ample evidence that chronic exposure to secondhand smoke increases the risk of cardiovascular disease, and therefore heart attack risk, and there is some suggestive evidence that acute exposure to secondhand smoke may present some degree of risk to individuals with existing severe coronary artery disease, there appears to be no scientific basis for claims that brief, acute, transient exposure to secondhand smoke increases heart attack risk in individuals without coronary disease, that it increases such risk to the level observed in smokers, that it can cause fatal or catastrophic cardiac arrhythmias, or that it represents any other significant acute

The rent-seekers' motivation for simplifying and distorting is not hard to see. The exaggeration of risks has the direct effect of creating greater public support for the policies they regard as best. It also has the indirect effect of making the cultural environment less hospitable to opposing groups, such as those who wish to smoke. This puts further pressure on individuals to stop smoking because they will find themselves uncomfortable in more and more public spaces. Thus the paternalist net can widen by increasing the number of those who, for self-interested or moralistic reasons, will support more inclusive bans.

Exaggeration of risks from overweightness and obesity. Distortion and simplification by rent-seekers is even more evident in the public debate on overweightness and obesity. Both the World Health Organization (WHO) and the National Institutes of Health (NIH) have chosen to define "overweight" as having a Body Mass Index (BMI) of 25-29, "obese" as having a BMI of 30 or more; but as J. Eric Oliver observes, there is little or no scientific basis for these definitions. Both organizations claim to have based their definitions on evidence of elevated mortality among people with BMI above 25, "but in both the WHO and NIH reports, none of the research really substantiated this claim."[128] For instance, the NIH report cited the work of nutritionist Richard Troiano for support, even though his work largely contradicts the claim of elevated mortality:

> Not only did [Troiano] discover that mortality was highest among the very thin as well as the very heavy, but also that the increased mortality was typically not evident until well beyond a BMI level of 30. And until one gets to a BMI of 40 or more, the differences in mortality are still within the bounds of statistical uncertainty.[129]

Earlier BMI-based thresholds of overweightness and obesity were not as high, and the current BMI-based definitions were not adopted until 1997, following recommendations by the WHO and NIH. Dropping the overweightness threshold to BMI of 25 added almost 40 million Americans to the group considered to be at risk. What drove the new recommendations? Oliver notes that most of the WHO report, which influenced the NIH's ultimate decision, "was drafted and written under the auspices of the International Obesity Task Force (IOTF)," an organization that "primarily funded by Hoffman-LaRoche (the maker of the weight-loss drug Xenical) and Abbott Laboratories (the maker of the weight-loss drug Meridia)."[130]

More broadly, public agencies, health researchers, and pharmaceutical firms all have a strong interest in cultivating the notion that obesity is a "disease," as doing so opens the doors to funding, research grants, and tax breaks. All three groups have participated in

---

cardiovascular health hazard in nonsmokers. *In light of this, the claims that are being widely disseminated by a large number of tobacco control groups appear to be scientifically unjustified and inaccurate*." Emphasis added.

[128] J. Eric Oliver, FAT POLITICS: THE REAL STORY BEHIND AMERICA'S OBESITY EPIDEMIC (2006), 22.

[129] Oliver (2006), 25.

[130] Oliver (2006), 28-29.

the effort to publicize, and exaggerate, the risks associated with overweightness and obesity.[131] The decision by Medicare in 2004 to classify obesity as a disease cleared the way for medical coverage of weight-loss drugs, diet programs, and bariatric surgery (as distinct from obesity-correlated illnesses such as diabetes, which were already covered).[132]

As the second-smoke and obesity examples suggest, rent-seekers with an interest in distorting and simplifying information come in at least two varieties. The first variety is old-style paternalists who believe they know best and don't necessarily care about the underlying preferences of the targets. Traditional temperance and health advocates fall within this category. They sacrifice the "public interest" in terms of the preferences of the targets to their own moralistic goals. The second variety is people who stand to benefit economically from the promotion or cessation of some activity. Examples include mutual fund companies that provide savings instruments, weight-loss clinics and programs, and manufacturers of smoking-cessation drugs. Public officials and agencies with an interest in preserving and expanding their domains also fall within this category, as do some individuals in their role as consumers and workers (for instance, non-smoking bar customers who would prefer to have more establishments cater to their tastes).

It is worth noting that rent-seeking activities impart a particular *direction* to slippery slopes. Whether motivated by their selfish economic interests or a moral agenda, rent-seekers cause asymmetrical movement along a gradient. The savings industry has an incentive to encourage marginally *higher* estimates of optimal savings rather than marginally lower estimates. Pharmaceutical companies that produce weight-loss drugs support marginally *lower* thresholds for overweightness and obesity. Smoking-cessation interests encourage marginal *extensions* in the public spaces in which smoking is prohibited. As a result, we are unlikely to observe "backward" slippage toward more laissez-faire policies.

c. Expanding Justification Slopes

An effective means by which a proponent can argue for a new policy is to show that the accepted justification for an existing policy also provides the foundation for a new one. This is rational in most circumstances because policymakers, especially voters, lack the ability, time, and energy to analyze each policy proposal on its own merits. They will be rationally ignorant about most proposals. Therefore, if a new policy is seen as a small extension of an existing policy, decisionmakers will tend to defer to the perceived rationales behind existing policies, or at least not strongly resist new policies based on them. This tactic is especially effective in the presence of a gradient, since that makes it easier to find policies whose distance from existing policy is sufficiently small. Volokh refers to this phenomenon as the "small change deference heuristic."[133]

---

[131] Oliver (2006), 47-51.
[132] Elizabeth Wolfe, *Medicare Redefines Obesity as an Illness*, Associated Press (July 16, 2004). Accessed March 14, 2008, at http://mpdu.com/newap/D83RS0N80.html.
[133] Volokh (2003), 1108.

Since a new policy is being advocated later and the old policy was adopted earlier, the rationale for the old policy must be reconstructed. Reconstruction is not simply a replication of the arguments historically produced at the time of the original discussions. Many different arguments doubtless will have been made. Rather, it is an *interpretation* or *rationalization* based on current understanding of the meaning or function of certain policies. It is this which is relevant for current decisions. A given law may function to promote a particular goal, whether or not that was in fact the goal that led to the law's passage.[134] If decisionmakers assume the existing law is (all things considered) justified, then a similar law with a similar function may be deemed a good idea too. Again, this assumption is natural in the context of rationally ignorant decisionmakers.

How does the reconstruction of the rationales behind older policies work? Slippery slope theorists have noted a tendency for complex principles to be simplified in the process of rationalizing existing rules and policies. Frederick Schauer refers to this tendency as the "bias in favor of simple principles." [135] Eric Lode quotes Justice Cardozo, who observes that "the half truths of one generation tend at time to perpetuate themselves in the law as the whole truth of another, when constant repetition brings it about that qualifications, taken once for granted, are disregarded or forgotten."[136] Volokh draws attention to the process of simplification in the policy realm: "Sometimes, the debate about a statute will focus on one justifying principle... But as time passes, the debates may be forgotten, and only the law itself will endure; and then advocates for future laws B may cite law A as endorsing quite a different justification."[137]

i. The Logic of Justificatory Expansion

We have already seen the simplification process at work in the previous subsection, when considering the tendency of experts to simplify their own theories. Now we will consider how simplification occurs more broadly, regardless of who does it. There are at least three possible patterns by which justifications expand in the process of reconstructing them.

> **Substituting broader for narrower justifications.** A relatively narrow justification J1 leads to the adoption of policy P1. Later, in reconstructing the origin of P1, observers conclude that some broader justification J2, which is sufficient to justify *both* P1 *and* P2 (the new policy proposal), was the real reason for P1's adoption. The authority conferred on J2 by the existence of P1 thus increases the likelihood of P2 also being adopted.

In the context of the new paternalism, an example of this phenomenon is provided by the substitution of outsiders' preferences for those of target agents. The *initial* (i.e., new

[134] It is important to recognize that the rationalization of past policies is not simply an intellectual exercise. In a context such as the above, it is part of a pragmatic argument designed to attain certain ends in the present.
[135] Schauer (1985), 372.
[136] Lode (1999), 1516.
[137] Volokh (2003), 1089.

paternalist) argument is that government can help advance the target agents' own "true" preferences. This argument, honestly applied, can only justify a limited set of policies. However, the initial justification is easily replaced with the simpler and broader argument that government can help advance the target agents' welfare according to some exogenous definition of the good. That justification obviously supports much greater intervention.

As we have seen, the experts themselves have already taken the first step in the simplification process by choosing between competing preferences of the target agent, without any basis for doing so. It is not hard to believe that non-experts will make the same simplification and take it even further. They might, for instance, follow Sunstein and Thaler's lead in seeing direct evidence of the "need" for intervention in the form of low savings rates, high obesity rates, and so forth. According to the sophisticated new paternalist justification, however, these are not sufficient to demonstrate suboptimal choices according to agents' subjective preferences; but we cannot assume that future voters, politicians, and bureaucrats will infer the sophisticated justification from existing policies.

> **Paring multiple justifications down to one**. The initial policy P1 is adopted based on the joint support of two justifications, J1 and J2. Later, in reconstructing the origin of P1, observers focus on the most obvious or salient justification J1 while ignoring the supporting role of J2. This increases the likelihood of adopting policy P2, which is supported by J1 but not J2.

Non-paternalistic justifications can interact with paternalistic ones to support policies that would not succeed with one justification alone. Earlier, we suggested that socialization of health costs creates a greater incentive to regulate lifestyle choices (such as smoking, overeating, and motorcycling without a helmet) on grounds of protecting the taxpayers' pocketbooks. This situation is not *strictly* paternalistic because the justification is the interests of others. However, such arguments can play a supporting role. A lifestyle restriction that both saves tax dollars *and* might induce better individual choices as well stands a greater chance of passage.[138]

The slippery slope risk arises from the possibility that, after their passage, lifestyle restrictions will be reinterpreted as having arisen largely or entirely from paternalist concerns. If so, then other policies – those that lack the buttressing justification of saving taxpayer dollars – naturally follow. While such policies might not have been supported initially, the *is-ought* heuristic can lead rationally ignorant policymakers to assume the paternalistic justification is solely (not just jointly) sufficient. If autonomy-based

[138] Amartya Sen, for example, weaves together harm-to-self arguments with non-paternalistic harm-to-others arguments in favor of restricting tobacco consumption. "Once acquired, the habit of smoking is hard to kick, and it can be asked, with some plausibility, whether youthful smokers hava an unqualified right to place their future selves in such bondage." He then indicates that the "victims of self-choice" should be considered among the "others" harmed by smoking. He goes on to augment this argument by claiming that public health expenditures and the associated taxes can justify restrictions on tobacco consumption. . *See* Sen (2007).

objections were not enough to block the existing policy, why should they block the newly proposed extension?

> **Stripping away qualifications on justifications**. An initial policy P1 is adopted with the support of justification J, with significant qualification Q (which specifies some circumstances where J would not apply). Later, in looking back, observers note J but fail to consider the importance of Q. This increases the likelihood of adopting policy P2, which would fail under J-with-Q, but is supported by J unlimited by Q.

We have already seen an example of this phenomenon, in Camerer and coauthors' criterion for asymmetric paternalism. That criterion indicated that if the ratio of irrational to rational people is high enough, in comparison to the ratio of policy costs (to rationals) and benefits (to irrationals), then the policy is justified. Camerer and coauthors summarized this criterion by saying a policy is most likely justified as long as there are *at least some* irrational people. In other words, an important qualification – that benefits be high relative to costs – had been stripped out.

ii. Application to Smoking Bans

The expansion of justifications does not take place in a vacuum. The activities of targeted agents, policymakers, experts, and rent-seekers interact to alter the terms of the public debate. In addition, justificatory expansion can interact with other slope processes. To show these interactions, we offer the case of smoking bans.

As discussed earlier, the debate on environmental tobacco smoke (ETS) has been affected by simplification and distortion of the evidence about the risks from secondhand smoking. The simplification of facts is, however, only a facet of a more complex process, as the justifications for restricting ETS have also expanded over time.

The original Surgeon General's report on the dangers of smoking[139] generated demand for various kinds of legislation to reduce the incidence of smoking, including higher cigarette taxes, health warnings, bans on tobacco advertisement on television. However, such direct paternalistic efforts had their political limits. Attention then turned increasingly to the dangers of ETS or secondhand smoke. As emphasized earlier, this is not a case of *pure* paternalism since the emphasis was mostly on the harm to others that results from "involuntary" exposure to tobacco smoke. Non-paternalistic justifications played the primary role in justifying further legislation, such as bans on smoking in public buildings, with paternalistic justifications playing at most a supporting role.

[139] U.S. Department of Health, Education, and Welfare, *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service* (1964).

It became increasingly clear, as further studies were published, that smoking restrictions in public spaces had the additional effect of reducing primary smoking.[140] These results could be regarded as a simply a "side-effect" of a policy whose primary goal was to curtail harm to others. But to the extent that this was seen as a benefit, the paternalistic aspect of the smoking bans became more prominent. As the political demand for smoking restrictions grew, it began to exceed the scientific basis for claims of unavoidable harm to bystanders.[141] Thus, more and more, the emphasis shifted to the effect of curtailing smoking itself, that is, to the paternalistic aspect of public-smoking restrictions. As Ronald Bayer and James Colgrove put it:

> But it was precisely because restrictions on public smoking had important effects on smoking itself that many public health activists gave such emphasis to broadening the range of prohibitions. It would have been impossible to ignore the fact that measures initially pursued in the name of protecting nonsmokers had secondary benefits – restricting smoking itself – that far outweighed the contribution associated with limiting exposure to ETS.[142]

For example, the CDC now "strongly" recommends public smoking bans on two grounds: the reduction of workplace exposure to ETS *and* the decrease in daily smoking or increased rates of cessation among smokers.[143] Going further, anti-smoking advocates have begun to emphasize reduced cigarette consumption almost exclusively, with little reference to protecting third parties:

> *More stringent* clean indoor air laws are associated with decreased smoking prevalence and cigarette consumption and a higher proportion of quitters. … *Comprehensive* public clean air laws have the potential to reduce prevalence and consumption rates of the entire population (including nonworkers and non-indoor-working smokers) by amount 10 percent. Additionally, clean air regulations may contribute to a changing social norm with regard to smoking by altering the perceived social acceptability of smoking. Because of changes in social attitudes and the need to smoke in less hospitable places, smokers may be induced to attempt to quit or not initiate.[144]

---

[140] Matt Barry, *Smoke-Free Laws Encourage Smokers to Quit and Discourage Youth from Starting*, Campaign for Tobacco-Free Kids Fact Sheet (July 27, 2006); Prabhat Jha, *Curbing the Epidemic: Governments and the Economics of Tobacco Control*, The World Bank (1999), 51-53.

[141] See supra 125-127 and accompanying text.

[142] Ronald Bayer & James Colgrove, Science, Politics, and Ideology in the Campaign Against Environmental Tobacco Smoke, 92 American Journal of Public Health 949 (2002), 953.

[143] Centers for Disease Control and Prevention (CDC), *Strategies for reducing Exposure to Environmental Tobacco Smoke: Increasing Tobacco-Use Cessation, and Reducing Initiation In Communities and Health-Care Systems: a Report on the Recommendations of the Task Force on Community Preventive Services* (2000), 5 ("Smoking bans, effective in reducing exposure to ETS, also can reduce daily tobacco smoke consumption for some tobacco users and help others quit entirely.") and 6, Table 2 ("Smoking bans and restrictions" "Strongly recommended"). Available at http://www.cdc.gov/mmwr/PDF/rr/rr4912.pdf, accessed on April 2, 2008.[143].

[144] Diana Bonta, *Clear Air Laws*, in Richard J. Bonnie, Kathleen Stratton & Robert B. Wallace (eds.), ENDING THE TOBACCO PROBLEM: A BLUEPRINT FOR THE NATION, B-1 (2007), B-2. Emphases added.

Thus, laws whose initial purpose was to protect bystanders were *reinterpreted* as a means of protecting people from themselves. Their benefits were measured, not in terms of fewer nonsmokers being diagnosed with smoking-related conditions, but in terms of less smoking overall.[145]

This history provides an example of the first simplification pattern described earlier, in which a policy initially justified in terms of one policy is later interpreted as having a different justification – one that can justify further interventions. Or we may see it as an example of the second pattern, in which multiple justifications are whittled down to one. Either way, the stage is set for further legislation that is difficult if not impossible to support based on the initial justificatory foundation. The recent proposals and legislation designed to ban outdoor smoking at, say, beaches have minimal health value for nonsmokers.[146] A further step in this direction is the growing movement to restrict smoking in apartment-style housing.[147] Although there is still some reference to the ETS-based justification, paternalistic concerns do the heavy lifting.

To summarize, the justification process for public smoking bans took us through the following steps:

1. The original antismoking paternalism was a direct effort to reduce tobacco smoking. This had political limits which encouraged anti-tobacco activists to turn to the issue of "involuntary" exposure to ETS.
2. Initially, indoor public space smoking bans were justified largely by the goal avoiding ETS harm to others, with purely paternalistic goals playing a secondary role.
3. Research revealed that these bans reduce primary smoking.
4. Rationalizations of the function of public smoking bans then focused on paternalistic goals regarding to the behavior of smokers, potential smokers who might be affected by the public example of others, and nonsmokers willing to tolerate public smoking.

---

[145] Gruber and Koszegi (2001), 1289, show that, on the assumption that smoking in different areas is complementary, if it cannot be regulated in the home, it should be "overregulated" outside in, say, restaurants and bars. This will cause a reduction in smoking in the home. Thus, the shift in justificatory emphasis noted above is a rational adjustment to data showing the interrelation of the two areas of smoking. If smoking is very inconvenient (costly) outside of the home some people may stop smoking altogether.

[146] *See*, for example, Jim Giles, Andy Coghlan, and Linda Gedes, *Anti-smoking groups accused of distorting the science on the risks of heart attack,* 196 NEW SCIENTIST 8 (November 10, 2007): "…Oakland [California] said last month that smokers could no longer light up at bus stops and ATMs. Down the coast in Calabasas, smoking in public places, including the street, has been illegal for over a year."

[147] Belmont, California has banned "smoking in apartment blocks and shared houses… In Maine… officials say that 40 percent of all shared residential buildings are now smoke-free." *See* Giles et al. (2007). This issue has now been raised with regard to co-op apartments in New York City. Some smoke from one apartment may seep into another apartment. One solution is to make ensure that the smoke pathways are sealed. Whose obligation it is to pay for this is contested. Should it prove expensive, a case could be made that smoking at home should be restricted. The justification will then have *both* paternalistic and harm-to-others aspects. *See* Bradley Hope, *Latest Hot Co-op Topic: Secondhand Smoke*, THE NEW YORK SUN 1 (December 6, 2007), 1, 13. Available at http://www.nysun.com/article/67569, accessed March 15, 2008.

5. Based on the expanded, and now largely paternalistic, function of *existing* public smoking bans, further restriction of public smoking became acceptable with little or no evidence of significant harm to bystanders.

Note that expanding justifications do not tell the whole story of anti-smoking regulation. At least two other processes contribute. First, the efforts of rent-seekers to distort the facts about ETS helped to increase public support for such regulation. Second, the combined effect of anti-smoking laws has been to "undercut the social support network for smoking by implicitly defining smoking as an antisocial act."[148] In addition, such laws have restricted the ability of nonsmokers to accommodate smokers by exposing themselves voluntarily to ETS that is not especially harmful. Consequently, an "attitude-altering slope"[149] has buttressed the expanding justification slope, as more nonsmokers now feel entitled to have smoke-free environments provided for them (even when the health risks are negligible).

d. On Experts vs. Regular People

Experts, and more broadly intellectuals like the readers of scientific and law journals, naturally respond to sophisticated argumentation. The complex interaction of multiple justifications is their favored milieu, the drawing of distinctions their stock in trade. Some of the claims of this part might, therefore, might seem anti-intellectual or unfair, because we are discussing the *misinterpretation* of the new paternalists' arguments, rather than the new paternalists' actual arguments. Why can't the experts simply reject the simplification, distortion, and expansion of their justifications for policy?

The answer is twofold. First, intellectuals cannot always control the development of their own ideas. Many regular people, whose job is not the careful parsing of sophisticated arguments, nevertheless affect the policy process. These regular people include voters, of course, but in varying degrees other public decisionmakers, such as politicians, bureaucrats, and some judges. The point is not that such people are stupid, but that they are rationally ignorant. They act based on simplified versions of arguments because they do not have the time, energy, or motivation to explore the sophisticated versions. In short, *simple is easy; complex is hard*.

Second, decisionmaking takes place in a social context. The fact that some people will recognize certain distinctions as relevant does not mean that others will. The decisionmakers who create a policy are not necessarily the people who enforce it, or who interpret it, or who consider extensions of it. We therefore need to keep in mind Bernard Williams's distinction between "reasonable distinctions" and "effective distinctions." The former are distinctions for which a reasoned argument can be made, whereas the latter are distinctions that can be defended "as a matter of social or psychological fact."[150]

---

[148] Stanton A. Glantz, *Achieving a Smokefree Society*, 76 CIRCULATION 746 (1986), 747.
[149] Volokh (2003), 1077ff.
[150] Quoted in Lode (1999), 1479.

The social and psychological facts, in a world of rational ignorance, often point toward simplification and even distortion of both theory and fact.

6. SLIPPERY SLOPES WITH COGNITIVELY BIASED POLICYMAKERS

In previous sections of this article, we have described slippery-slope processes generated by the new paternalism on the assumption that policymakers behave rationally.[151] In this section we drop that assumption. We now assume that the policymakers are no different from the targeted individuals. They too exhibit cognitive and behavioral biases. We make this assumption for the sake of argument. New paternalists who justify their policies on grounds of cognitive and behavioral biases of targets must confront the issue of biased policymakers, including voters. We argue that such biases would reinforce the slippery-slope phenomena we have been discussing.

This section is the most tentative for three reasons. First, the cognitive limitations we discuss have not been tested in the particular slippery-slope contexts addressed here. Second, we have not undertaken a systematic analysis of every specific bias that has been identified by behavioral science.[152] This would prove to be a very difficult task, since the effects of these biases often run in opposite directions, have different degrees of importance, and interact with each other. Third, it is not clear in any given situation that policymakers will have the *same* biases as targeted agents. It is conceivable that they could have offsetting biases.

Nevertheless, we have chosen to analyze what we believe are the most general and widespread cognitive limitations. Most of these were chosen for emphasis by Daniel Kahneman in the revised version of his Nobel lecture.[153] In other words, we have not cherry-picked our biases.

a. Hyperbolic Discounting by Policymakers

Policymakers can have short time horizons for various reasons. They might no longer hold office when future costs and benefits of their policies occur. Insofar as voters have imperfect memories, they might fail to fault policymakers for the ill effects (or credit them with the good effects) of policies they supported. Both of these effects give fully rational policymakers an incentive to discount future consequences.

---

[151] We also assumed that the experts and rent-seekers are rational.

[152] There are just too many. *See* Joachim I. Krueger and David C. Funder, *Towards a Balanced Social Psychology: Causes, Consequences and Cures for the Problem-Seeking Approach to Social Behavior and Cognition*, 27 BEHAVIORAL AND BRAIN SCIENCES 313 (2004), 317, Table 1. The authors present forty-two errors or biases identified since 1985; they consider it a partial list.

[153] Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics* 93 AMERICAN ECON. REV. 1449 (2003).

If policymakers are hyperbolic discounters, there is yet another reason they will tend to discount the future: because they apply *especially* high rates of discount when some costs or benefits are in the present (or near future).

If so, then just as regular people may succumb to temptations like desserts and cigarettes that promise short-term pleasures, we should expect policymakers to succumb to "policy temptations" that generate short-term political gains. For instance, they might be tempted in election years to adopt policies, such as fiscal stimulus bills and trade restrictions, that will improve their electoral chances while pushing costs into the future.

How does this worsen slippery-slope risks? Slippery slope events are necessarily sequences that play out over time: policy A's adoption now leads to policy B's adoption later, leading to policy C's adoption yet further in the future. Hyperbolic discounting implies that when policymakers are faced with a policy proposal that is appealing in the present, but which creates a danger of bad policies being adopted further down the line, they will be inclined to focus on the former at the expense of the latter. In short, they will be less cognizant of slippery-slope risks.[154]

For instance, policymakers might be tempted to create a small fat tax, on grounds that it will induce marginally "better" eating decisions. Opponents might argue that adopting a small fat tax will create a danger of a larger fat tax in the future, as future policymakers – having already incurred the costs of creating a tax collection mechanism – see the opportunity to increase their tax revenues and fund special-interest constituencies. If they are hyperbolic discounters, the policymakers will not take this risk seriously enough, even if they recognize it as real.

Like hyperbolic discounters in the private sector, policymakers should be expected to exhibit time inconsistency: the tendency to make commitments and promises and then break them when the moment of choice arrives. They might, for instance, repeatedly express a willingness to take measures to fight budget deficits in the future, while nevertheless passing bloated budgets and incurring large debts in the present. In the slippery-slope context, note that critics of slippery-slope arguments will sometimes claim to be able to resist the urge to adopt bad policies in the future. The idea is that we can "do the right thing today, and resist doing the wrong thing tomorrow."[155] They might, for instance, promise to keep fat taxes relatively low (and linked to scientific evidence about the extent of present-bias). The existence of time inconsistency bears directly on the plausibility of promises to do the right thing in the future even in the face of temptation.

As we have indicated, hyperbolic discounting does not imply that far-sighted preferences are necessarily the best. Thus, we might be accused of inconsistency in saying the

---

[154] *See* Herbert Spencer, *The Coming Slavery*, in THE MAN VERSUS THE STATE 31 (1981 [1884]), 43: "But the 'practical' politician who, in spite of such experiences repeated generated after generation, goes on thinking only of proximate results, naturally never thinks of results still more remote, still more general, and still more important than those just exemplified."

[155] Volokh (2003), 1029-1030: "The slippery slope argument, opponents suggest, is the claim that 'we ought not make a sound decision today, for fear of having to draw a sound distinction tomorrow.'"

myopic behavior of policymakers is necessarily bad. In response, we direct attention to the new paternalists' own assumptions in this regard. They assume, without justification, that long-run discount rates are the appropriate standard. If so, then targeted agents do indeed have a problem in making wise choices, but policymakers with the same biases will have a similar problem – and one that will affect all citizens, not just those most in need of correction. On the other hand, the new paternalists are free to adopt our position that there is no objectively correct rate of discount, but in that case they must abandon their claim that targeted agents are in need of correction.

We have also argued that people afflicted by excessive impatience have various self-debiasing mechanisms at their disposal, such as imposing internal rewards and punishments, structuring their external environment, and enlisting the help of third parties (like families and support groups). Policymakers may have access to similar devices. We suggest that the analogous devices in the policy arena usually take the form of institutional constraints, such as judicial review and constitutional limitations on what areas can be regulated by government. The greater need for external restraints follows from the fact that bad self-governance by a single person primarily affects that person, whereas bad governance by policymakers affects all of the governed. Thus, the individual has a rational incentive to rein in his own irrational impulses, whereas policymakers' incentive to do so is attenuated. In other words, policymakers are more likely to exhibit "rational irrationality"[156] about matters of personal choice than are the private citizens who make those choices.

b. Narrow Framing by Policymakers

Hyperbolic discounting may be regarded as a specific variety of narrow framing, that is, the tendency to focus on particular aspects of a decision problem rather than seeing overall (including long-run) consequences. Narrow framing arises, in large part, because immediate and concrete effects are more psychologically accessible than remote and abstract ones. Adam Gifford notes the relationship between hyperbolic discounting and the accessibility of the concrete:

> Similar problems, however, can arise when making choices between two goods when both are available to the agent with a predetermined identical short delay. If one of the two goods is represented only by a printed word, for example, and the other good is visible to the agent when making the choice, reversing the level of abstraction of the two goods can result in a reversal of the agent's choice.[157]

The concreteness of specific problems – such as a perceived low savings rate, the readily observed expansion of waistlines, the rising costs of healthcare, and so forth – will tend to focus policymakers on immediate policy choices without regard for more distant consequences. As Daniel Kahneman puts it, in the context of purely private decisionmaking, "The problem at hand and the immediate consequences of the choice

---

[156] Bryan Caplan, THE MYTH OF THE RATIONAL VOTER (2007).
[157] Adam Gifford, Emotion and Self-Control, 49 Journal of Economic Behavior and Organization 113 (2002), 114.

will be far more accessible than other considerations, and as a result problems will be framed far more narrowly that the rational model assumes."[158] There is no reason to think policymakers can counter this problem any better than private citizens.

Narrow framing leads decisionmakers to consider choice-options simply as they arise, framed by present circumstances, the crisis of the moment, and perhaps that activities of rent-seekers. Their actions will often be *ad hoc* solutions to particular problems, and the narrow framing produces a tendency not to see important interrelationships. In Kahneman's words again, "The decision of whether or not to accept a gamble is normally considered as a response to a single opportunity, not as an occasion to apply a global policy."[159] For example, the interaction of biases may be ignored. This means the problem is not simply one of discounting long-term effects, but also of discounting effects that occur through longer and more complex chains of causality.

Narrow framing will enhance every variety of slope we have discussed so far, because all slopes occur in part from a failure to take a global perspective on policy. Altered incentives slopes, for instance, occur because policymakers tend to focus on one issue at a time – in this case, a single cognitive or behavioral bias, or a single means of correcting a bias. Simplification and distortion slopes occur because policymakers enact policies to address a specific problem, while failing to see how the new policy could empower experts and rent-seekers to advance less desirable policies in the future. To the extent that narrow framing inhibits policymakers' awareness of such possibilities, it exacerbates the slippery slope risk.

Moreover, as we have emphasized repeatedly, slippery slopes are most likely in the presence of a gradient. Movement along gradients is assisted by a narrow focus on two very close or similar cases, where the second case is considered to have a certain characteristic such as acceptability simply because the first did as well. A gradient slope is easier to resist when policymakers consider the whole spectrum of possibilities rather two similar policies or cases at a time.

c. Passive Framing Accepted by Policymakers

As presented in the behavioral literature, framing does not result from the deliberate choices of the decisionmaker; instead, it is an aspect of decision problems that is *passively accepted*.[160] It is the result of unconscious processes whereby the conscious mind sees options or events with particular features accentuated; framing alters "the relative salience of different aspect of the problem."[161]

Here we suggest that the particular way in which the new paternalists (most notably Camerer and coauthors and Sunstein and Thaler) have framed the issue of paternalism

[158] Kahneman (2003), 1460.
[159] Kahneman (2003), 1459.
[160] Kahneman (2003), 1459: "The basic principle of framing is the passive acceptance of the formulation given."
[161] Kahneman (2003), 1450.

gives rise to an inherently expansionary dynamic. If irrational or boundedly-rational policymakers accept the new paternalists' approach, they will have accepted a paternalism-generating framework. Thus future policymakers, or the same policymakers in future situations, will tend to see more opportunities for paternalistic intervention than they otherwise would.

The decisions of targets are not intrinsically different from those of the policymakers. Framing is thus important in the policy context as well. The *public-policy* framework produced by the new paternalists directs policymakers' attention to intrapersonal preference conflicts, that is, conflicts between operative preferences (choosing the sugary dessert) and deeper or more important preferences (maintaining good health). The framework then labels as paternalism any plan that alters the decision problem with the intent of improving welfare. Therefore, if there is to be any solution to the target's problem, paternalism is inevitable. Thus, the decision problem is framed not as "whether or not paternalism is desirable," but as "what form of paternalism shall we have." Sunstein and Thaler, for example, urge us to "abandon the less interesting question of whether to be paternalistic or not, and turn to the more constructive question of how to choose among the possible choice-influencing options."[162]

To the extent that policymakers accept the new paternalists' framing of the problem of intrapersonal preference conflict, they shall be led to produce all manner of paternalistic schemes. Yet it is far from necessary to look at matters in this way. An alternative perspective, which steers clear of the new paternalist framework of trying to extricate people exogeneously from decision contexts in which they cannot promote their own welfare, is available.[163]

People subject to temptation assign utility not only to individual options, but also to decision problems or sets of options that they will face in the future. Thus, a person may assign a higher utility to a decision problem that omits a certain tempting, but ultimately welfare-reducing, alternative. As a result – and as we have emphasized earlier – they will deliberately make choices to structure their decision contexts. A vegetarian who still likes the taste of meat may prefer not having meat on the menu of a restaurant. Such a person will spend real resources to constrain his future options or to commit himself to a certain option in advance. A dieter may raise the cost of the less-preferred option by placing cookies on a high shelf where they can be gotten only with increased effort. People may announce to their friends that they will lose weight thus suffering embarrassment if they do not. They may also wager money on their own self-control.[164] Other market participants may offer these people controlled options to make a profit. This could include cafeterias with no unhealthful sweets, or with the sweets are placed in

---

[162] Sunstein & Thaler (2003), 1166.

[163] See, for example, Faruk Gul and Wolfgang Pesendorfer, *Temptation and Self-Control*, 69 ECONOMETRICA 1403 (2001), and Faruk Gul and Wolfgang Pesendorfer, *Self-Control and the Theory of Consumption* 72 ECONOMETRICA 119 (2004).

[164] A new web-based business, stickK (http://www.stickk.com/about.php), allows people to sign "commitment contracts" to motivate themselves to achieve goals, with self-selected consequences of failure. The inspiration for the site came, in part, from research in behavioral economics.

less tempting locations, or with especially good fruit alternatives. Recently, manufacturers have sold portion-controlled bags of cookies or chips to help individuals with self-control problems.[165] With something akin to this framework in mind, paternalistic options would come to the fore only *later*, if it is clear that the costs of self-control or pre-commitment or market choice-restriction, including discovery of the targets' preferred decision problems and enforcement costs, are higher than the costs of paternalistic intervention.[166]

In this alternative framework, the policymaker or analyst is first led to search for ways in which individuals or markets might control their choice sets. Then, if he sees no such methods in operation he will go back and decide whether in fact the apparent preference for an omitted option is true or merely cheap talk. People may fail to construct restricted preference sets because they have no real interest in restraining their choices. Only when the analyst is satisfied that the issue is the prohibitive costs of private control can the discussion of government paternalism begin.

Therefore, the Sunstein and Thaler approach is expansive not only in the sense that adoption of specific policies today will make the adoption of further, even more interventionist, policies more likely in the future, but also because their basic framework of analysis frames the overall issue as one in which some form of paternalism is "inevitable." Sunstein and Thaler adopt a *paternalism-generating public-policy framework*. If policymakers accept this framework, they will be led by the framing to produce more and more paternalistic policies.

d. Extremeness Aversion

Framing effects can, according to some work in behavioral economics[167], result in violations of the preference axiom of "independence of irrelevant alternatives." In other words, when choosing between option X and Y, some decisionmakers will choose X over Y when Z is not option, but then choose Y over X when Z is available. Somehow the mere presence of Z in the chooser's mind alters his attitude about the trade-off between X and Y. We suggest that such effects – specifically, the extremeness aversion identified

---

[165] Jeremy W. Peters, *In Small Packages, Fewer Calories and More Profit*, THE NEW YORK TIMES (July 7th, 2007). Accessed March 16, 2008, at http://www.nytimes.com/2007/07/07/business/07snack.html.

[166] To be more precise, paternalistic intervention would be considered only when its costs are lower than the lowest of the following: self-control, self-commitment, and market choice-restriction costs.

[167] Itamar Simonson & Amos Tversky, *Choice in Context: Tradeoff Contrast and Extremeness Aversion* 29 JOURNAL OF MARKETING RESEARCH 281 (1992); Joel Huber, John W. Payne, and Christopher Puto, *Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis*, 9 JOURNAL OF CONSUMER RESEARCH 90 (1982); John W. Payne, James R. Bettman, and Eric J. Johnson, *Behavioral Decision Research: A Constructive Processing Perspective*, 43 ANNUAL REVIEW OF PSYCHOLOGY 87 (1992); Itamar Simonson, *Choice Based on Reasons: The Case of Attraction and Compromise Effects*, 16 JOURNAL OF CONSUMER RESEARCH 158 (1989).

by Itamar Simonson and Amos Tversky – can increase the likelihood of initial passage of new paternalist policies and ease the movement along the slope to harder paternalism.

Extremeness aversion refers to the finding that "the attractiveness of an option is enhanced if it is [presented as] an intermediate option in the choice set and is diminished if it is [presented as] an extreme option."[168] For instance, Simonson and Tversky found that experimental subjects choosing between a low-end camera and a medium-quality camera split about equally between the two options – but when they were also presented with a high-end camera, they became substantially more likely to choose the medium-quality camera.[169]

The new paternalists' rhetorical device of positioning their proposals as the "middle ground" exploits, perhaps inadvertently, the phenomenon of extremeness aversion . In advocating their favored "soft" paternalism, they make a point of introducing "hard" paternalism for contrast. Camerer and coauthors, for instance, say:

> For those (particularly economists) prone to rigid antipaternalism, the paper describes a possibly attractive rationale for paternalism as well as a careful, cautious, and disciplined approach. For those prone to give unabashed support for paternalistic policies based on behavioral economics, this paper argues that more discipline is needed and proposes a possible criterion.[170]

Similarly, Jolls and Sunstein say their approach "adopts a middle ground between inaction or naïve informational strategies, on the one hand, and the 'insulating' strategies of heightened liability standards or outright bans, on the other."[171] Thus, in characterizing the choice between two policies, absence of paternalism and their own proposals, the new paternalists make a point of introducing a third option that should – in theory – be irrelevant to the choice between the first two.

More formally, we may characterize the alternatives in terms of the tradeoff between autonomy and welfare. "Anti-paternalism" is characterized by high autonomy and low welfare, as shown by point AP on Figure 1. At the other extreme is "hard paternalism," (HP), characterized by low autonomy and high welfare.[172] "Soft paternalism" (SP) obviously lies somewhere in between.

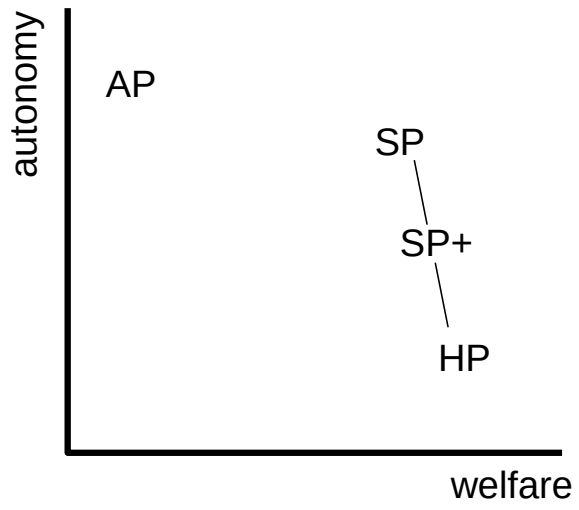
If we suppose that policymakers, too, are affected by extremeness aversion, that tendency can be exploited to move policy further down the slope to harder paternalism. The mechanism works through the introduction of another policy – let us call it "soft paternalism plus" (SP+) – that is intermediate between the newly adopted soft paternalist policy (SP) and hard paternalism (HP), as illustrated in Figure 1. The latter two options

---

[168] Simonson & Tversky (1992), 281.
[169] Simonson & Tversky (1992), 290.
[170] Camerer et al. (2003), 1212-1213.
[171] Jolls & Sunstein (2006), 16.
[172] We do not agree that hard paternalism will in fact lead to higher welfare; this is, however, how the tradeoff seems to have been characterized by the new paternalists themselves.

now constitute the extremes of comparison, making those policymakers affected by extremeness aversion more favorably disposed toward the middle ground of SP+. If SP+ is adopted, that becomes a new endpoint when another intermediate policy, SP++, is introduced. The middle ground continues to move toward the paternalistic extreme. [173]

Figure 1: Extremeness Aversion in Policy Choice



To illustrate concretely, the new paternalist might begin with the claim that people are not saving enough for retirement and make the policy recommendation for mandated automatic 401K enrollment for employees who do not explicitly opt out. This, he argues, will increase savings. A more intrusive paternalist policy (which he does not currently advocate) might be a compulsory increase in the savings rate – say, through a payroll-tax deduction. The new paternalist thus argues, "See how moderate I am – neither the extreme advocate of laissez-faire nor the hard paternalist of mandatory savings."

If this argument and policy are accepted, then the left edge of the continuum will move. The new "laissez-faire" extreme is now simply the mandate for a default with an exit option. The policy middle-ground will have shifted. The slightly harder new paternalist may argue for a further step in the world of defaults – this time for something like the "Save More Tomorrow" (SMarT) program which *automatically increases the employees' rate of savings* with each increase in salary.[174] This too could be mandated as a default – again, the employee could opt out. Now the "laissez-faire" edge is the SMarT default. A

[173] It is also worth noting the connection of extremeness aversion to gradients. The process described here works only so long as policies can be found that are intermediate between the present policy, whatever that might be, and harder paternalism. When policy exists on a spectrum, it is always possible to find such policies. If a given "intermediate" policy is rejected as too far from the status quo, then another policy closer to the status quo can be introduced.

[174] Shlomo Benartzi and Richard H. Thaler, *Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving*, 112 J. POLITICAL ECONOMY S164 (2004).

yet harder paternalist could now argue that requiring 401K enrollment, with full freedom to choose investment options *but no opt-out possibility* is a desirable middle ground. If accepted, this becomes the new "laissez-faire" position. From there, especially if people choose investment vehicles that are too risky, restriction of these options may seem to be a middle-ground policy. At each step of the process, the middle ground shifts.

Of course, this has not yet occurred. Another example, in this case a progression that has actually occurred, can be found in the expansion of anti-smoking restrictions. If we think of public-smoking restrictions as occupying a continuum of costs to smokers, we can see how the middle-ground moves by steps along this continuum. Initially, when smoking was restricted in only minor ways or in few places, the costs to smokers of the restrictions were relatively small.

One of the first smoking restrictions, mandated by the Civil Aeronautics Board in 1973, was simply to separate smokers and nonsmokers in separate sections on airplanes. As a middle ground between an outright smoking ban and laissez-faire, this seemed a reasonable low-cost policy option. Now this became the "laissez-faire" position. Further costs were imposed upon smokers by Federal Aviation Administration (FAA) in 1988 which banned smoking on all scheduled domestic flights under two hours in duration. Two hours of abstention was a middle-ground between the mere separation of passengers and the higher costs imposed by abstention for many hours. This middle-ground position imposed somewhat larger costs that most smokers could still presumably tolerate. With that as the new left end of the spectrum, the FAA in 1990 banned smoking on all scheduled domestic flights. In 2000, the U.S. Department of Transportation extended the ban to all U.S. international flights. The airplane ban was complete. But not all airports are smoke-free yet.[175]

The new paternalists offer a policy framework that emphasizes the middle ground, not just in specific policy areas such as smoking regulation, but as a general perspective. Yet the middle ground is not a stable place; what constitutes the middle ground is a function of what policies have already been adopted. To the extent that the new paternalists' middle-ground argument is successful and policymakers adopt policies on this basis, there is a potentially powerful dynamic at work.

d. Prototype and Affect Heuristics: Extension Neglect

The behavioral literature finds quantitative or "extension neglect" in many different contexts.[176] For example, when asked to determine their willingness to pay for wildlife preservation or other public goods, respondents tend to ignore quantitative dimensions; in

[175] The history is recounted in Bonta (2007), B-6.
[176] Kahneman (2003), 1463-1466.

one experiment, subjects' willingness to pay to save migratory birds from drowning in an oil pond was more or less the same for 2,000 or 20,000 or 200,000 birds saved.[177]

The root of the phenomenon, according to Kahneman, is the *prototype heuristic*. A prototype is a representative instance of the class. It may be an average of particular features or simply what comes to the agent's mind, perhaps as a result of recent or spectacular instances. In the example described above, the prototype might be the image of a single bird dying in a pool of oil; the image does not account for the *number* of birds that might be spared this fate.

For our purposes the most important feature of a prototype is that it is extensionless. A prototype is an *exemplar*; it is not the set of all instances. The size of the set can grow without the exemplar changing, and thus an increase in the extensional (that is, quantitative) aspect of the class is neglected in the prototype.[178] Yet it is the prototype that affects actual decisionmaking. Once the prototype is formed, the *affect heuristic* enters the picture. The valuation of the entire class will be determined by the attitude or affective response of the agent to the prototype.[179]

i. Extension Neglect in the New Paternalist Framework

If policymakers are also subject to extension neglect, as a result of using prototypes in the evaluations of policy, they will be more susceptible to slippery slopes. The most likely prototypes of Sunstein and Thaler's libertarian paternalism are (a) the deliberately placement of fruit before sugary desserts in cafeterias, and (b) the adoption of default enrollment of employees in 401K programs. In the first case the cost of opting out by customers is extremely low; in the second case employees can opt out at only slightly higher costs. Sunstein and Thaler have worked hard to make these the prototypes.[180] Their case for paternalism always begins with one of these cases, before gradually extending to cases that are less purely "libertarian." As discussed earlier, the analysis moves from the prototypical (low cost-of-exit) cases, to legal presumptions that can be

---

[177] Kahneman (2003), 1463-1464, citing William H. Desvousges et al., *Measuring Natural Resources Damages with Contingent Valuation: Tests of Validity and Reliability*, in Jerry A. Hausman, ed., CONTINGENT VALUATION: A CRITICAL ASSESSMENT (1993), 91. Another example of extension neglect isd duration neglect. When asked to rate the degree of discomfort of a colonoscopy people will ignore the length of time the procedure takes, concentrating only on the "representative" moments of peak and end. Donald A. Redelmeier & Daniel Kahneman, *Patients' Memories of Painful Medical Treatments: Real-time and Retrospective Evaluations of Two Minimally Invasive Procedures*, 66 PAIN 3 (1996).

[178] Kahneman (2003), 1464: "Other things equal, an increase in the extension of a category [e.g., the number of birds in the above example] will increase the value of its extensional attributes [the total number of birds saved], but leave unchanged the values of the prototype attributes [the value of the representative bird].") Bracketed remarks supplied.

[179] Daniel Kahneman, Ilana Ritov, and David Schkade, *Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues*, 19 JOURNAL OF RISK AND UNCERTAINTY 203 (1999), 206: "The attitude to a set of similar objects is often determined by the affective valuation of a prototypical member of that set."

[180] Thaler & Sunstein (2003), Sunstein & Thaler (2003). See also the Wall Street Journal debate between Richard Thaler and Mario Rizzo, *Econoblog: Should Policies Nudge People to Make Certain Choices?* (May 25, 2007), available at http://online.wsj.com/public/article/SB117977357721809835-45dCZESztTYwbcmLpVZEpaSe790_20070531.html.

waived only under general conditions set by the state, to working-hours limitations that can be waived only for time-and-a-half overtime, to mandatory cooling-off periods for consumer purchases that cannot be waived at all.[181] The progression proceeds in terms of greater costs of opting-out by the putative beneficiaries. In addition, and little noticed by Sunstein and Thaler, the progression can also proceed in terms of greater costs imposed on employers, sellers, and other parties.

In our earlier discussion of the gradients created and exploited by the new paternalism, we found it reasonable to blame information costs for policymakers' potential insensitivity to incremental changes. The introduction of prototype analysis and extension neglect, however, increases the likelihood that these quantitative or extensional features may be ignored entirely. Policymakers will tend to focus on the prototype (the relatively easy cafeteria placement or 401K issues) and their affective response to further policies will be determined in part by their putatively positive response to the prototype. Therefore, approval of the initial, least intrusive policies will increase the likelihood that more intrusive policies will be adopted through the interplay of the prototype and affect heuristics.

In general, the new paternalists' paradigm is one that trusts policymakers to make careful, nuanced calculations of costs and benefits[182] – calculations that depend on both the extent and degree of cognitive biases. One example of the need for measures of the extent of bias is provided by Camerer and coauthors' "asymmetric paternalism" criterion. The criterion's application to specific policies depends crucially on a quantitative variable, the *fraction* of people considered irrational. A better version of the criterion would also account for heterogeneity in the *degree* of irrationality. This, too, is a quantitative variable. We may expect policymakers to ignore both of these – assuming they are subject to the same biases that behavioral economists attribute to the rest of us.

ii. Extension Neglect in the Calculation of Optimal Sin Taxes

The taxation of immediate-gratification goods provides a more specific example of how extension neglect may enhance slippery slope processes. It has been suggested that a tax on "junk food" or cigarettes might eliminate or reduce the "internalities" that present selves with little willpower impose upon future selves.[183] Today a person may enjoy junk food or cigarettes but, in so doing, he may create future costs that far exceed the present

[181] See discussion in section 4b.

[182] Sunstein & Thaler (2003), 1166: "First, programs should be designed using a type of welfare analysis, one in which a serious attempt is made to measure the costs and benefits of outcomes… Second, some results from the psychology of decisionmaking should be used to provide ex ante guidelines to support reasonable judgments about when consumers and workers will gain most by increasing options."

[183] The term 'internalities' was coined in Richard J. Herrnstein et al., Utility Maximization and Melioration: Internalities in Individual Choice, 6 Journal of Behavioral Decision Making 149 (1993), 151, defining it as "a within-person externality … which occurs when a person underweighs or ignores a consequence of his or her own behavior for him- or herself." The concept of internalities is employed by various authors in the new paternalist literature, including Gruber and Koszegi (2001), O'Donoghue and Rabin (2003a, 2003b), and Camerer et al. (2003), 1221, note 50. For a critical analysis, see Whitman (2006).

benefits. A carefully calibrated internality tax would impose upon the individual today a cost sufficiently high to make him take these future consequences into consideration.

From the new paternalist perspective the issue is not merely reducing consumption of immediate gratification goods, but reducing it to an *optimal level*, that is, to a level consistent with the present benefits of consumption just exceeding future costs, where the costs have been discounted appropriately.[184] So in the new paternalist equilibrium, people will still eat junk and smoke cigarettes, but to an optimal degree.

If we take this policy approach seriously, then the policymaker must base his consumption tax on two pieces of information. The first is the rational degree of impatience needed to determine what "excessive" impatience is. The second is the distribution of rates of excessive impatience across consumers; not everyone has the same rate, because some people have greater willpower than others. The optimal tax should balance the benefits from reducing the consumption of those with relatively less willpower against the costs of reducing the consumption of those with relatively more willpower. Too high a tax can be worse than too low a tax.

How does a non-expert policymaker deal with these quantitative complexities? If he uses the heuristics discussed in this section, his approach will largely be to ignore them. Perhaps the most psychologically available pictures are the prototypes of a compulsive cigarette smoker or obese junk-food consumer – that is, persons with little willpower who seek immediate gratification and endanger their future health. *How many* such people there are, and *how much* correction each of them needs, will likely have little influence on policymaking. Prototypes, by definition, contain no reference to the distribution of attributes within the relevant class. The affect or evaluation evoked by the prototype will be, presumably, a generalized disapproval of junk-food consumption and smoking with little or nothing in the way of quantitative distinctions.

Suppose an expert produces a policy paper with specific estimates relating to impatience, willpower, and optimal tax rates, and *mirabile dictu*, it is adopted by the policymakers. Will matters stand there? This is unlikely. To the extent that policymakers, especially voters, deal with the issue through prototype and affect heuristics, with their resulting quantitative neglect, constraints to limit further taxes are quite loose. The reasoning follows a simple path: "People smoke and eat junk food, thereby endangering their health. Health is good; that which endangers it is bad. Perhaps they should be taxed until they stop." This is counter to the new paternalists' theory, but the new paternalists do not control the policy process. There is no effective bar to sliding beyond their modest proposals.

7. REJOINDERS & RECOMMENDATIONS

a. Rejoinders to New Paternalist Responses

[184] Keep in mind that new paternalists are *not* claiming that the proper rate of discount is zero.

In their book *Nudge*, Sunstein and Thaler recognize the slippery-slope objections to their policies, and they offer three responses. We reply to their responses here.

Sunstein and Thaler's first response is that the slippery-slope argument "ducks the question of whether our proposals have merit in and of themselves" They say if the initial interventions are worthwhile, then we should "make progress on those, and do whatever it takes to pour sand on the slope…"[185]

Our claim is not that slippery slopes are the *only* objection to the new paternalism. Various other objections have also been made (and referenced in the introduction to this article). The slippery slope is an *additional* argument against the new paternalism.

The idea that we should "make progress" on the initial interventions, and then do what we can to "pour sand" on the slope, is a variant of the usual (and, we think, hackneyed) response to all slippery slope arguments: that we can simply "do the right thing now, and resist doing the wrong thing later." But if the slope argument is correct, there is a causal (albeit probabilistic) connection between initial interventions and later ones. Saying we should move forward on those initial interventions is akin to saying we should do something because it promises present benefits, while ignoring the potential costs in the future. (Ironically, it is just this sort of error in private decisionmaking that most new paternalists think cries out for correction.) *The slope risk must be counted among the costs of the initial intervention*.

Furthermore, how should we "pour sand" on the slope? Aside from invoking the term "libertarian" (see below), Sunstein and Thaler offer no suggestions. We do, in the remainder of this article. Our suggestions involve, among other things, rejecting their paternalist-generating framework.

Sunstein and Thaler's second response is that their own "libertarian condition" limits the steepness of the slope. They say their proposals are "emphatically designed to retain freedom of choice."[186]

In short, they are relying on the "libertarian" part of libertarian paternalism to do the work of resisting paternalism. But as we have seen (see especially section 4b), their redefinition of "libertarian" actually encourages the slope. They recognize no sharp line between libertarian and unlibertarian policies, just a smooth gradient. And also as we have seen, their proposals do *not*, in fact, preserve freedom of choice in all cases. They have proposed or supported numerous policies (such as mandatory time-and-a-half overtime pay and cooling-off periods) that rule out certain options altogether, all under the rubric of libertarian paternalism.

It is also simply implausible to think the mere word "libertarian" will create a bulwark against further interventions. Even if Sunstein and Thaler themselves genuinely care

[185] Sunstein & Thaler (2008), 237.
[186] Sunstein & Thaler (2008), 237.

about freedom of choice, they cannot control the application and transformation of their own ideas. They will not be in charge of all future legislation. As we have emphasized throughout this paper (see especially sections 3 and 5), the creation of policy is a social process that involves multiple decisionmakers, who may not share their alleged concern with freedom of choice.

Sunstein and Thaler's third response is to insist that in many situations, "some kind of nudge is inevitable," because there will always be default rules and contexts that frame choices in certain ways.[187]

It is one thing to have defaults, quite another to choose them with paternalist goals in mind.[188] Traditional contract law chooses defaults in line with the customary expectations of the parties in question. Thus the new paternalists advocate overruling customary expectations so as to the privilege what they (the experts) believe are better decisions. They would purposely shift transaction costs to those who wish to deviate from the experts' preferred outcomes.

If new paternalism were truly inevitable, it would hardly be necessary to argue for it. Clearly, Sunstein and Thaler believe they are offering something beyond the inevitable. Moreover, they present their position in a manner designed to ease the transition from the inevitable to the more intrusive. They explicitly reject any sharp line between changing defaults and raising costs in other ways. Again, their very own next step, in discussing default rules, is to suggest *raising the cost of exercising exit options*, and then to endorse *eliminating some options altogether*.[189]

b. Minimizing the Danger

How, then, might we protect ourselves against paternalist slopes? We have three recommendations, addressed both to the new paternalists themselves and to those who might be persuaded by them. These recommendations are intended to lower the probability of adopting new paternalist policies to begin with, but also to help resist more intrusive policies after initial policies have been adopted.

Have reasonable expectations of decisionmakers. One lesson of behavioral economics is that we cannot reasonably expect decisionmakers to carefully consider the full ramifications of their choices in light of the best available evidence. Instead, they economize on information by using choice heuristics, and they sometimes myopically focus on present and concrete problems while ignoring more distant and abstract ones. This is no less true of public decisionmakers (including voters, politicians, judges, bureaucrats, experts, and rent-seekers) than it is of private citizens. Indeed, the problem is likely worse for public decisionmakers, because they lack the incentives to discover

[187] Sunstein & Thaler (2008), 237.
[188] Mitchell (2005), 1259-60: "But not all default rules are equally paternalistic or paternalistic in the same way, and … it does not follow from libertarian principles that the central planner should choose the default rule that enhances the welfare of affected individuals."
[189] See section 4b.

and control their own cognitive limitations. Private decisionmakers at least face the costs and benefits of their own mistakes, and thus have an incentive to correct them.

It is therefore insufficient to ask policymakers to carefully weigh the costs and benefits of each new paternalist proposal. The "careful, cautious, and disciplined approach" advocated by Camerer and coauthors[190] is rather unlikely to guide real-world policy. We should not expect policymakers to weigh all the economic, scientific, and psychological evidence objectively, to stand on nuanced distinctions, and to adopt policies that carefully target just those people who need help most. We should expect policies to be blunt instruments.

Reject the paternalist-generating framework. The new paternalists say that the framing of problems can affect decisions by emphasizing certain aspects of a situation and downplaying others. As we have argued earlier (see especially section 6c), the new paternalists themselves have framed the public policy debate in a manner that emphasizes opportunities for intervention while downplaying or ignoring private alternatives. Adopting that framework increases our vulnerability to slippery slopes.

In contrast to this paternalist-generating framework, we recommend a slope-resisting framework – one that emphasizes the limitations of public policy and the potential for private solutions. In this alternative framework, both private and public decisionmakers are understood as having essentially the same cognitive defects; they also have various tools for self-correction. For private decisionmakers, the tools include resolutions and commitments, conscious construction of their environment, and voluntary submission to social controls from third parties. For public decisionmakers, the tools include procedural, substantive, and attitudinal limitations on the scope and extent of government action.

Maintain important distinctions. Slippery slopes, including paternalist ones, can sometimes be resisted by standing on easily enforceable bright-line rules. One such bright-line rule is the distinction between public and private decisionmaking. Another is the distinction between coercive and non-coercive intervention. John Stuart Mill enunciated these distinctions in terms of the Harm Principle, which says restriction of individual choice is justified only on grounds of harm to others. He argued:

> [The individual] cannot rightfully be compelled to do or forbear, because in the opinions of others, to do so would be wise or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise.[191]

Of course, Mill understood, and we agree, that if a person harms himself and in so doing violates his legal responsibilities to others, he ought to face the relevant legal penalties.[192]

---

[190] Camerer et al. (2003), 1212.
[191] John Stuart Mill, ON LIBERTY (Stefan Collini, ed.) (1989 [1859]), 13.
[192] Mill (1989 [1859]), 81.

But the State has no legitimate interest *that can be vindicated by coercion* strictly in the prevention of harm to oneself. To the extent that policy adheres to this principle, the paternalist slope will obviously never get started.

We do not contend that a single violation of Mill's Harm Principle will send us hurtling toward heavy-handed paternalism in all areas of life; if that were true, we would already be doomed. We do contend that increasing numbers of such interventions, passed under the guise of helping people do better by their own preferences, and without any recognition of the lines being crossed, will tend to create momentum toward further interventions. Keeping the Harm Principle clearly in mind – and recognizing any given restriction on autonomy (however small) for what it is – will, we hope, retard movement down the slope.

To some extent – especially in Sunstein and Thaler's book *Nudge* and Daniel Ariely's book *Predictably Irrational* – the new paternalists have presented their position as self-help advice. That is, they offer behavioral economic insights into achieving better self-control and personal management. Under the Harm Principle, such efforts are perfectly unobjectionable. They fall in Mill's category of remonstration, reasoning, persuasion, and entreaty.

The problem, as we have argued, is that the new paternalists do not clearly distinguish private, voluntary efforts from public, mandatory ones. Instead, they deliberately construct a continuum from soft to hard paternalism (see, especially, section 4b). They define freedom of choice in terms of the cost of exercising a given option, without regard to whether the costs are imposed coercively or by the voluntary choice of resource owners. In this way, they effectively erase a reasonably bright-line rule – the distinction between private action and state coercion – and purposely replace it with a gradient. They also regularly present public and private, and coercive and non-coercive, paternalistic activities alongside each other, without recognizing any important distinction between them, and often simply ignoring the coercive aspects of their policies (e.g., the way in which allegedly pro-employee policies limit the freedom of the employer).

We suspect the new paternalists resist bright-line rules and encourage gradients because of an unavoidable feature of rules: they nearly always err by both over- and under-inclusion. A rule that allows private paternalism but not public paternalism would admit some varieties of paternalism that new paternalists might oppose, such as Wal-Mart's restrictions on what sort of movies it will stock; and it would disallow some varieties of paternalism they favor, such as mandatory terms in employment contracts. That is, however, the price of having rules. The compensating advantage of rules (or at least one advantage) is providing a bulwark against the problems of vagueness, including the threat of slippery slopes.

Bernard Williams's distinction between logical and effective distinctions is frustrating, because it means we cannot always rely on the normative distinctions that make most sense to us. But it is also enabling, because it reveals that some distinctions may be

useful – that is, effective – without being strictly logical. Thus, even if the new paternalists do not think the public-private and coercive-non-coercive distinctions track their ideal notions of right and wrong, such distinctions might nevertheless be practical as guides for law and policy.