PATERNALIST SLOPES

1/23/07 draft copy

Please do not quote without permission.

Douglas Glen Whitman

and

Mario J. Rizzo[1]

Abstract

A growing literature in law and public policy harnesses research in behavioral economics to justify a new form of paternalism. Contributors to this literature typically emphasize the modest, non-intrusive character of their proposals. A distinct literature in law and public policy analyzes the validity of "slippery slope" arguments. Contributors to this literature have identified various mechanisms and processes by which slippery slopes operate, as well as the circumstances in which the threat of such slopes is greatest.

The present article sits at the nexus of the new paternalist literature and the slippery slopes literature. We argue that the new paternalism exhibits many characteristics identified by the slopes literature as conducive to slippery slopes. Specifically, the new paternalism exhibits considerable theoretical and empirical vagueness, making it vulnerable to slopes resulting from altered economic incentives, enforcement needs, deference to perceived authority, bias toward simple principles, and reframing of the status quo. These slope processes are especially likely when decisionmakers are subject to cognitive biases – as the new paternalists insist they are. Consequently, soft paternalism can pave the way for harder paternalism. We conclude that policymaking based on new paternalist reasoning should be considered with greater trepidation than its advocates have suggested.

Keywords: paternalism, slippery slopes, behavioral economics

JEL Classifications: B5, D7, I0, K0

[1] Department of Economics, California State University, Northridge, and Department of Economics, New York University, respectively. Glen.Whitman@csun.edu and Mario.Rizzo@nyu.edu

A growing literature in law and public policy harnesses research in behavioral economics to justify a new form of paternalism.[2] The thrust of the argument is straightforward: Human beings are not fully rational (in the sense traditionally used in economic theory), but in fact exhibit an array of cognitive problems, including but not limited to status quo bias, optimism bias, hindsight bias, context dependence, susceptibility to framing effects, and lack of willpower. These cognitive problems lead to errors in decision making, meaning that people systematically behave in ways that fail to advance their own best interest. Insofar as actual behavior deviates from optimal behavior, governments (as well as other people and institutions) can potentially intervene in ways that will improve the individual's well-being.

The leading contributors to the "new paternalist" literature, as we shall call it, place great emphasis on the modesty of their proposals. The policies advocated are said to be minor and non-intrusive. A recent feature article in *The Economist* captures the tenor:

> Their aim is not the 'nanny state', a scold and killjoy forcing its charges to eat their vegetables and take their medicine. Instead they offer a vision of what you might call the 'avuncular state,' worldly-wise, offering a nudge in the right direction, perhaps pulling strings on your behalf without your even noticing.[3]

---

[2] Colin Camerer, Samuel Issacharoff, George Lowenstein, Ted O'Donoghue, and Mathew Rabin, Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism,' 151 University of Pennsylvania Law Review 1211 (2003); Richard H. Thaler and 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(4) University of Chicago Law Review 1159 (2003); Christine Jolls and Cass R. Sunstein, Debiasing Through Law, 35 Journal of Legal Studies 199 (2006); Jonathan Gruber and Botond Koszegi, Is Addiction 'Rational'? Theory and Evidence. 116 *Quarterly Journal of Economics* 1261 (2001); Ted O'Donoghue and Matthew Rabin, Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes, 93 *AEA Papers & Proceedings* 186 (2003); Ted O'Donoghue and Matthew Rabin, Optimal Sin Taxes, unpublished manuscript, Cornell University, University of California at Berkeley (2003).

[3] The avuncular state, *The Economist*, April 8th-14th 2006, 67-69, 67.

Christine Jolls and Cass Sunstein, for instance, repeatedly refer to their proposals for debiasing behavior through law as a "middle ground" between laissez-faire and more heavy-handed paternalism[4], one that is a "less intrusive, more direct, and more democratic response to the problem of bounded rationality."[5] Colin Camerer, et al., characterize their "asymmetric paternalism" model as a "a careful, cautious, and disciplined approach" to evaluating paternalistic policies.[6] In general, the new "soft" paternalism is presented as a kinder, gentler form of paternalism that avoids the problems of the older "hard" paternalism.

A distinct literature in law and public policy analyzes the validity of "slippery slope" arguments.[7] A slippery slope argument is one suggesting that a proposed policy or course of action that might appear desirable now, when taken in isolation, is in fact undesirable (or less desirable) because it increases the likelihood of undesirable policies being adopted in the future. Despite the poor reputation of slippery slopes as a form of argument, recent work by various authors has rehabilitated slippery slope reasoning by identifying the specific mechanisms and processes by which slippery slopes operate, as well as the circumstances in which the threat of such slopes is greatest.

The present article sits at the nexus of the new paternalist literature and the slippery slopes literature. The new paternalist approach exhibits many of the characteristics identified in the slopes literature as conducive to the occurrence of

---

[4] Jolls and Sunstein 2006, 208, 216.
[5] Ibid., 201.
[6] Camerer, et al., 2003, 1212.
[7] 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).

slippery slopes. Indeed, new paternalist policies, and the theories that support them, are *permeated* by these dangerous features. As a result, soft paternalism – even if initially modest and non-intrusive – has the potential to pave the way for harder paternalism, including some policies of which the new paternalists themselves would disapprove. We conclude that policymaking based on new paternalist reasoning ought to be considered with much greater trepidation than its advocates have suggested.

In Part I, we offer a brief defense of slippery slope reasoning, in general and as applied to the new paternalism. In Part II, we discuss the primary factor that makes the new paternalism especially vulnerable to slippery slopes: theoretical and empirical vagueness, which create a natural gradient between softer and harder paternalist policies. In Part III, we apply several specific slope processes (or mechanisms) to new paternalist policymaking. The specific processes include altered economic incentives, enforcement needs, the *ad verecundiam* heuristic (i.e., deference to perceived authority), bias toward simple principles, and reframing of the status quo. Finally, in Part IV, we briefly discuss the implications of slippery slope risks for evaluating policy proposals.

Part I. A Defense of Slippery Slope Reasoning

Although the slippery slope literature does not speak with a single voice, we think the general conclusion is clear: most careful analysts have concluded that, while slippery slope arguments are not universally valid, they cannot simply be dismissed. Some slope arguments are valid and others are not. The key to distinguishing them is to identify the

specific processes or mechanisms by which slopes occur, as well as the circumstances that affect the likelihood of such slopes.[8]

Nevertheless, slippery slope arguments suffer from a poor reputation. As Eric Lode notes, the slippery slope has even been classified as a fallacy in many introductory logic texts.[9] A short defense therefore seems in order.[10] The most common response to the slippery slope argument is that it immediately crumbles in the face of any logical or reasonable distinction between the (presumably good) policy under consideration and the (presumably bad) policy to which it will allegedly lead. "We can do the right thing now," the response goes, "and resist doing the wrong thing later." The main problem with this reply is that it trades on an ambiguity in the word "we." The present decisionmaker and the future decisionmaker need not be the same. Even if present decisionmakers are willing and able to make the relevant distinctions, future decisionmakers may be unable or unwilling to do so. The proponent of a slippery slope argument need not show that policy A logically entails policy B, only that adoption of A increases the likelihood of future decisionmakers adopting B – even if doing so would be illogical or mistaken.

Put somewhat differently, we ought to heed Bernard Williams's distinction between "reasonable distinctions" and "effective distinctions." Reasonable distinctions are those for which one can make a sensible argument, whereas effective distinctions can be defended "as a matter of social or psychological fact."[11] These need not be the same; some reasonable distinctions will not be honored in practice, while some arbitrary (non-reasonable) distinctions can be successfully defended. The critic of slippery slope

[8] See, especially, Volokh 2003, Lode 1999, and Rizzo and Whitman 2003.
[9] Lode 1999, 1474.
[10] For a short defense of slippery slopes in the context of a different policy debate, see Eugene Volokh, Same-Sex Marriage and Slippery Slopes, 33 Hofsta L. Rev. 1155, 1163-1165 (2005).
[11] Quoted in Lode 1999, 1479.

argumentation focuses on the existence of reasonable distinctions – but effective distinctions are the ones that matter.

Moreover, slippery slope arguments are especially apropos in addressing the new paternalism. Our approach here might seem unfair, inasmuch as we are criticizing the new paternalists *not* primarily for the actual positions they have advocated[12], but for the unwarranted positions that ignorant or illogical people may draw from them. Recall, however, that the new paternalists' arguments rely on the existence of just such ignorant and illogical people. New paternalist policies are justified precisely on grounds that many people have cognitive and behavioral biases that lead them to make systematic errors in their decisions. And as Eugene Volokh has argued, slippery slopes are closely connected to phenomena such as "bounded rationality, rational ignorance, [and] irrational choice behaviors such as context-dependence"[13]; this connection will become more apparent as the article proceeds. Thus, we suggest that the new paternalists' own arguments should drive them to fear the slope – perhaps even more than we do.

Furthermore, at least some new paternalists invite slippery slope arguments. Camerer, et al. do so explicitly: "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."[14] That is what we aim to do.

---

[12] We do that elsewhere; see Mario J. Rizzo and Douglas Glen Whitman, "Meet the New Boss, Same As the Old Boss: An Inquiry Into the New Paternalism," unpublished manuscript, New York University and California State University, Northridge (2006); Glen Whitman, "Against the New Paternalism: Internalities and the Economics of Self-Control," Cato Institute Policy Analysis no. 563 (2006).
[13] Volokh 2003, 1035.
[14] Camerer et al., 2003, 1251.

Part II. Gradients and Paternalism

A. Gradients as fertile ground for slippery slopes

Slippery slopes thrive in the presence of a continuum created by vague words or concepts, a phenomenon recognized by various slippery-slope analysts.[15] When words and concepts have fuzzy boundaries, it becomes difficult to defend sharp distinctions. Each case differs from the next case by only a small increment, so that unlike cases can be linked by a series of cases that differ only by degree. The classic example is the *sorites* paradox, named after the Greek word for "heap." How many grains of sand does it take to make a heap? If we already have a heap of sand and remove one grain, presumably we still have a heap. And the same is true if we remove another, and another… Repeatedly applying the premise that a heap minus one grain is still a heap, we eventually conclude that a single grain is a heap. That is a paradox, but not *merely* a paradox; it illustrates the difficulty of drawing lines in the presence of a gradient. In legal and policy contexts, the line-drawing dilemma can emerge whenever vague words or concepts are employed to define rules or the exceptions to them. Where is the line between mentally able and retarded (for purposes of capital punishment)? Where is the line between reasonable and unreasonable force (in defense of property)?

The presence of a vague term does not guarantee a slippery slope, but it increases the likelihood. The best defense against a slope – as the leading critique of slope reasoning implies – is the possibility of finding a clear (logical or practical) distinction among cases. Lacking such a distinction, decisionmakers will find it tempting to decide

[15] See especially Rizzo and Whitman 2003, 557-560; Volokh 2003, 1105-1114; Lode 1993, 1477ff.

new cases or adopt new policies on grounds of their similarity to existing cases and policies. Analogical reasoning economizes on information-gathering and calculation, allowing the decisionmaker to decide more quickly and with less effort. Note that this approach will be most appealing to boundedly rational decisionmakers – who, as the new paternalists emphasize, are common. The danger is that a chain of analogical reasoning can lead from sound to unsound decisions.

Lode argues that judicial decisionmaking is relatively more susceptible than legislative or bureaucratic decisionmaking to slippery slopes risks created by vagueness, and we are inclined to agree. The vulnerability of judicial decisionmaking to slopes results from the prevalence of analogical and precedent-based reasoning, as well as the tendency of judges "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."[16] But we think legislative and bureaucratic decisionmaking can also be vulnerable, for slightly different reasons.

First, legislators will sometimes purposely pass laws with vague language in order to finesse disagreements and avoid making tough decisions. The resulting laws will have to be interpreted by judges or administrative agencies (and their associated administrative courts).[17] Jolls and Sunstein, in discussing the modesty of their proposals in contrast to more intrusive legislation, draw attention to the existence of consumer protection laws that give administrative agencies a choice between requiring product information or

[16] Lode 1999, 1494.

[17] See Gary C. Bryner, Bureaucratic Discretion: Law and Policy in Federal Regulatory Agencies, 7 (1987): "Most 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. The responsibilities that have been delegated to them often greatly exceed the provided resources, thus necessitating important administrative choices and setting of priorities. *Some laws provide competing objectives that give administrators broad latitude*." (Emphasis added.)

banning the product outright.[18] So even if legislatures are capable of drawing sharp (perhaps arbitrary) lines to prevent sliding, that does not mean they will.

Second, legislatures can be affected by the lobbying pressure of groups with an interest in further legislation in a given area. Such groups can exploit the existence of a gradient to seek incremental changes that will largely go unnoticed by less organized groups. For example, financial services firms will have an interest in the expansion of default or mandatory savings schemes, as well as in affecting the policy particulars (e.g., what kinds of savings plans are eligible?). But the special interests involved need not be financially motivated, as there exist more "traditional" paternalist groups that have always favored more intrusive laws – for instance, religious groups that favor greater restriction of personal choice for moralistic reasons.[19] Another example is the Center for Science in the Public Interest, which advocates legislation to induce more healthful choices, with little hint of the new paternalists' recognition that other values (such as sheer enjoyment) might outweigh health concerns for some individuals.[20]

Third, gradients create fertile ground for legislative change when policy changes can affect the attitudes of voters and legislators – a claim that we will explain further in Part II. *Ad verecundiam* heuristics (i.e., deference to perceived authority), bias toward simple principles, and reframing of the status quo are all processes that can alter political attitudes, thereby making a slide down a gradient more likely.

---

[18] Jolls and Sunstein 2006, 207-8.

[19] 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."

[20] Jacob Sullum, "The Anti-Pleasure Principle," Reason Magazine, July 2003, downloaded from http://www.reason.com/issues/show/381.html on January 13th, 2007.

As Rizzo and Whitman note, vagueness in terms can arise from vagueness in the *theories* used to justify rules and policies, as well from vagueness in the *empirical application* of those theories.[21] It is in these respects that the new paternalist literature is most troubling.

B. Theoretical vagueness and hyperbolic discounting

Various paternalist policies have been justified by citing the notion of hyperbolic discounting. Traditional economic theory assumes that people's rate of trade-off or discounting between successive time periods is constant; that is, the trade-off between benefits at time T1 and at time T2 depends only on their distance from each other, not on their distance from the present. This is known as exponential discounting. But real people have inconsistent rates of discount: they exhibit higher rates of discount between time periods the closer those periods are to the present. This is known as hyperbolic discounting.[22] The result is that people exhibit time inconsistency: they will make decisions about future trade-offs, and then reverse those decisions later.

Hyperbolic discounting is used to explain self-control problems. Intuitively, people's inconsistent behavior reflects their vulnerability to temptation when those temptations are near. This creates a bias toward getting benefits now and incurring costs later: people spend too much and save too little, they consume too much and exercise too little, and so on. New paternalists have proposed various policies to deal with such self-control problems. Some have advocated automatic enrollment of employees in savings

[21] Rizzo and Whitman 2003, 574-578.
[22] The seminal article in this literature is Richard H. Strotz, Myopia and Inconsistency in Dynamic Utility Maximization, 23 Review of Economic Studies 165 (1955/56); see also George Ainslie, Breakdown of Will, Cambridge University Press (2001).

plans.[23] Others have advocated sin taxes, including fat taxes, as a means of inducing people to "internalize" the costs of their present behavior to their future selves.[24]

The theory of hyperbolic discounting, when used as a normative justification for policies to encourage greater self-control, involves considerable vagueness. While individuals may exhibit inconsistent rates of time discounting, there is no clear answer to the question of which rate of discount is the *correct* one. The new paternalists have typically assumed that the longer-term rate of discount is the appropriate one, but this assumption has no basis in theory. The behavioral inconsistency could be "fixed" to resemble exponential discounting (which generates no inconsistencies) by forcing individuals' short-term rate of discount to equal their long-term rate; but it could also be "fixed" by making the long-term rate of discount equal to the short-term rate.[25]

The new paternalist might reply that even if favoring the long-term perspective is arbitrary, it is not vague – it is a clear and obvious choice. But that clarity is an illusion created by the simplistic dichotomy between "short-term" and "long-term." The illusion is magnified by behavioral economists' frequent use of the *quasi-hyperbolic* time discount function, which represent an agent's short-term bias by means of a single parameter that gives extra weight only to the present. A quasi-hyperbolic discounter only has two rates of discount, the present rate and the future rate. As George-Marios Angeletos, et al., admit, the quasi-hyperbolic model "has been adopted as a research tool because of its analytical tractability"[26] – not because of its accuracy. In reality, people

[23] Thaler and Sunstein 2003, Sunstein and Thaler 2003.
[24] Gruber and Koszegi 2003, O'Donoghue and Rabin 2003a and 2003b.
[25] Whitman 2006, 5, 15, notes 17, 18.
[26] George-Marios Angeletos, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg, The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Estimation, 15(3) Journal of Economic Perspectives 47, 50 (2001).

exhibit true hyperbolic discounting, which means they display a *range* of different discount rates. For sufficiently distant choices, they may display no time discounting at all. There is thus no *single* future discount rate to favor by means of policy.[27] The decisionmaker who would implement policies to "fix" agents' intertemporal choices has to choose from a spectrum of possibilities, not just two. We can easily imagine decisionmakers sliding along the spectrum, initially enforcing only modest degrees of patience (say, with low fat taxes and low mandatory savings rates) and later shifting to higher and higher degrees of patience.

C. Theoretical vagueness and the correction of context-dependence

For some types of decision, people are subject to framing effects: one presentation of a decision problem will lead them to choose A over B, while another (logically equivalent) presentation of the same problem will lead them to choose B over A. One example of a framing effect is that medical patients will be more inclined to assent to a treatment described as having a 90% survival rate than one described as having a 10% death rate.[28] People also exhibit status-quo bias, a tendency to favor whatever is (or is presented as) the status quo or initial baseline situation.[29] An example is the persistent difference between willingness-to-accept (WTA) and willingness-to-pay (WTP)[30] – that is, the tendency of people to demand more money to part with an item than what they would pay to acquire the very same item, even when the item's value is low enough that it could create no significant wealth effects. Framing and status-quo bias are both forms of context-dependence – the tendency of people's decisions to change

[27] This follows from the form of the generalized hyperbolic discount functions most commonly employed in the psychology literature; see Angeletos, et al., 2001, 50.
[28] Sunstein and Thaler 2003, 1161, 1179.
[29] Russell Korobkin, The Status Quo Bias and Contract Default Rules, 83 Cornell L. Rev. 608 (1998).
[30] Sunstein and Thaler 2003, 1177.

depending on seemingly irrelevant aspects of the decision contexts. Some paternalist policies have been justified by the existence of context-dependence. Sunstein and Thaler, for instance, argue for the creation of new default rules in employment contracts, such as a presumed right to be fired only "for cause" rather than at will.[31] While it would remain possible to write contracts that override the default, and thus the same options as before would remain open, the new default would reframe the context to induce "better" choices (specifically, making employees more likely to reject "at will" employment).

The main theoretical difficulty with context-dependence as a justification for paternalist policy is similar to that of hyperbolic discounting: it relies on an internal inconsistency of an individual's preferences, but it gives no particular reason for favoring one preference over the other. The fact that someone has a higher WTA than WTP tells us that her attitudes are not consistent, but it does not tell us which figure is the correct one. The fact that a patient will assent to a medical procedure under description 1 but not under description 2 points up an inconsistency, but it does not tell us whether the medical procedure is worth doing – that would depend on preferences and attitudes toward risk.

Sunstein and Thaler emphasize that when people's choices are subject to context-dependence, the very meaning of "preferences" is unclear. "These contextual influences render the very meaning of the term 'preferences' unclear,"[32] they say; and "[i]f the arrangement of alternatives has a significant effect on the selections the customers make, then their true 'preferences' do not formally exist."[33] If there can be no appeal to true underlying preferences as the basis for favoring one frame of reference over another, then

[31] Sunstein and Thaler 2003, 1187.
[32] Sunstein and Thaler 2003, 1161.
[33] Sunstein and Thaler 2003, 1164.

some other external standard must be employed. Sunstein and Thaler do not specify the appropriate standard; instead they say, "We are not attempting to say anything controversial about welfare, or to take sides in reasonable disputes about how to understand that term."[34] But the standard of value chosen is the very essence of the problem. The justification for deliberate reframing of decisions to induce "better" choices therefore rests on a gaping theoretical lacuna. Different decisionmakers will naturally approach the problem with widely varying notions of welfare and well-being.

Does this theoretical vagueness create a gradient with slippery-slope potential? We believe it does. Although proposals like Sunstein and Thaler's genuflect to the notion of preserving individual choice, the underlying theory does not necessarily place any weight on choice. For any given standard of value, much more heavy-handed policies might be justified. The question, then, is how much weight the social welfare function ought to place on individual choice, and that parameter is not clearly specified by theory. There is no particular reason to think subsequent decisionmakers will rely on choice to the same extent as present ones in making their policy decisions. Given that individual choice plays no salient role in selecting the appropriate framing of decision problems, a gradient connects soft to hard paternalist policies. Policies that do not restrict individual choice differ from policies that mildly restrict individual choice only differ by degree – a point that Sunstein and Thaler recognize explicitly when they say, "[I]n all cases, a real question is the cost of exercising choice, and here *there is a continuum rather than a sharp dichotomy*."[35] Thus, statutes or judicial precedents that create freely waivable default rules lay the theoretical groundwork for default rules that can only be

[34] Sunstein and Thaler 2003, 1163, note 17.
[35] Sunstein and Thaler 2003, 1185; emphasis added.

waived at a cost, which in turn can lay the groundwork for default rules that cannot be waived at all.

D. Theoretical vagueness and context-dependence as a corrective device

Setting aside context-dependence as a justification for paternalist policy, some authors have suggested the use of context-dependence as a tool to solve problems created by other cognitive biases. Jolls and Sunstein cite research showing that consumers' optimism bias causes them to underestimate the risk of adverse consequences of certain products and services[36], and then suggest using the availability heuristic to address the problem. The availability heuristic is another variety of context-dependence in which the images and narratives presented with a decision problem affect the choices made, despite no objective difference in the facts of the situation. Jolls and Sunstein propose to make use of availability like so:

> 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.[37]

Put simply, firms would have to provide their customers with frightening stories to emphasize the seriousness of certain types of risk. But there is considerable vagueness about how frightening the narratives should be. Jolls and Sunstein are suggesting a switch from a bright-line rule (did the firm truthfully disclose the risk?) to a gradient standard (did the firm provide sufficiently scary examples?). They admit that showing customers worst-case scenarios can be counterproductive[38], which means there must be a

[36] Jolls and Sunstein 2006, 204-205.
[37] Jolls and Sunstein 2006, 212.
[38] Jolls and Sunstein 2006, 214.

means of distinguishing too-frightening from not-frightening-enough. "Of course there are line-drawing problems here," they say, "but the basic point is straightforward."[39]

In the presence of a slippery slope risk, line-drawing problems are of the essence, and neither the theory of optimism bias nor the theory of availability heuristics provides any clear guidance. There is no objective means, in practice or in theory, to distinguish between (a) customers who absorbed the relevant information and decided rationally to assume the risks and (b) customers who did not hear a compelling enough narrative about the risk. We can expect judges to decide new cases arising under "insufficient narratives" claims to make decisions by analogy with prior cases. Hindsight bias could play a role in making such decisions: given that an accident did occur, is it not obvious that the narrative was insufficient? The slope goes from missing narrative to mildly compelling narrative to worst-case-scenario narrative.

And does a narrative even have to be truthful? Jolls and Sunstein's policy description specifies a "truthful account of consequences," but nothing in theory requires that. Indeed, Sunstein and Thaler note the potential harm arising from some truthful information: "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."[40] Could a service provider (say, an HMO) be faulted for presenting such information? Once we have moved away from the notion of truthful information as the standard for liability, the appropriateness of any information (or lack thereof) depends entirely on the *result* in terms of consumer behavior. But again, mere results cannot tell us how to distinguish between (a) rational

[39] Jolls and Sunstein 2006, 214.
[40] Sunstein and Thaler 2003, 1183.

assumption or avoidance of risk and (b) behavior based on inadequate information about risk. *There is no objective standard for the "right" framing of a decision problem*.

And if it is sometimes appropriate to withhold information, might it not also be appropriate to misrepresent information – that is, to lie? Once more, the theory provides no reason to draw a line here. There is a gradient leading from merely providing information to reframing information to hiding information to providing deliberately incorrect information.

E. Empirical vagueness

Suppose, for argument's sake, that the new paternalist theories present no problems of theoretical vagueness: we have a theoretically valid means of selecting among intertemporal discount rates, of choosing among different framings of decision problems, and so on. Even so, the making of actual decisions and policies can run into a problem of empirical vagueness, meaning "indeterminacy in the application of a theory, typically created by lack of knowledge on the part of agents and decisionmakers who are expected to apply it."[41]

Consider policies designed to deal with hyperbolic discounting. Even supposing there exists a correct rate of discount, that does not mean decisionmakers will know it. The correct rate will presumably differ from person to person, and possibly from situation to situation (undersaving or overeating?). In addition, different people will respond to corrective policies in different ways; some will exhibit the desired response to the policy, while others might cut back on their own self-corrective efforts, while yet others might be too strongly affected by the policy. All of these factors are relevant for deriving the optimal policy devices to make people act on the correct discount rate. As we argue

[41] Rizzo and Whitman 2003, 577.

more extensively elsewhere[42], the informational requirements for choosing optimal debiasing policies are virtually insurmountable. Lacking the relevant information, decisionmakers will have to rely on incomplete research, guesswork, and – most troubling in the present context – reasoning by analogy. What is the appropriate size of a fat tax? What is the right amount to require people to save (or have saved by default)? The answers to these questions are *empirically* vague; we have insufficient knowledge to give precise answers.

Mathematical modeling can create the illusion of precision. A closed mathematical model can generate precise decision rules, defined in terms of all parameters included in the model. Calibrating the model to match reality is another matter entirely, particularly since a closed model necessarily excludes some potentially relevant variables. Consider, for example, Camerer, et al.'s criterion for good "asymmetric paternalism"[43]: 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 to exclude implementation costs and profits to firms.) This criterion seems clear enough in theory (though we might ask troublesome questions about the theory of value that generates B and C, especially in the absence of well-defined preferences). But the problem is in the application. How shall B and C be measured? What fraction of the public is subject to the form of irrationality in question? Moreover, as Camerer, et al. would surely admit, the model excludes any heterogeneity.

[42] Rizzo and Whitman 2006.
[43] Camerer, et al., 2003, 1219.

Everyone is either rational or not (no degrees of rationality), and everyone in either group gets the same benefit or harm. So what we have is, at best, a rule of thumb that is open to interpretation by specific decisionmakers – whether legislators, bureaucrats, voters, or judges.

In the context of their proposal to debias consumers via frightening narratives, Jolls and Sunstein admit that "the ultimate question of the optimal form of debiasing through the availability heuristic is an empirical one."[44] We have argued that important theoretical questions remain, but set aside that objection; there is still a matter of how to measure the appropriateness of framing. We lack a scale on which to measure fright, and we lack the knowledge to derive the right point on the scale. The answer will depend on the product or service in question, as well as the characteristics and personal histories of diverse consumers (what is frightening to me could be mundane to you). The specter of empirical vagueness looms large, and decisionmakers forced to decide in its presence will tend to rely on their own heuristics, including analogical reasoning. As suggested in the context of theoretical vagueness, hindsight bias could play a role here: when the one clear fact in the instant case is that someone was harmed by a product, it seems natural to place substantial weight on that fact alone.

To summarize, new paternalist proposals typically rely on models that are beset by theoretical vagueness, and that have the potential to create empirical vagueness. Vagueness makes the boundaries of key concepts fuzzy, creating gradients that connect good policies to bad, modest interventions to more intrusive ones. Decisionmakers who wish to economize on conceptual processing (in the presence of theoretical vagueness) and information processing (in the presence of empirical vagueness) will instead rely on

[44] Jolls and Sunstein 2003, 213.

other means of making decisions on new cases and policies. Those other means could easily involve the same cognitive biases and sources of error that the new paternalists have identified in regular people.

Part III. Applied Slippery-Slope Processes

A. Altered Economic Incentives Slopes

Slippery slopes can occur when the implementation of a new policy changes economic incentives (and thus behavior) in a way that makes other policies appear more desirable.[45] One simple example, offered by Rizzo and Whitman, is the effect that socialized medicine could have on regulation of lifestyle choices. To the extent that lifestyle choices (such as smoking, drinking, or risky sexual behavior) can increase healthcare costs, taxpayers under socialized medicine might be more inclined to support restrictions on lifestyle choices than they would under a system in which people bear (most of) their own health costs.[46]

New paternalist policies have the potential to alter economic incentives in ways that encourage further interventions in the future. We offer three examples:

The second-best problem. The second-best problem in economics refers to the fact that some market imperfections can, partially or totally, offset the effects of other market imperfections. As a result, correcting one imperfection without correcting another can actually exacerbate a problem.[47] For example, monopoly power will tend to

[45] Rizzo and Whitman 2003, 560-563.
[46] Rizzo and Whitman 2003, 556, 562.
[47] See, generally, Richard G. Lipsey and Kelvin Lancaster, A General Theory of the Second Best, 24 Review of Economic Studies 11 (1956).

increase the price of a good – which in general is undesirable. But what if production of the good involves negative externalities? In that case, policies that reduce monopoly power could result in more production of the good and thus greater pollution.

Douglas Besharov[48] demonstrates that a related problem applies *within* a person subject to cognitive biases: some biases can partially or completely compensate for others. As a result, attempts to fix one source of cognitive error can exacerbate others. For instance, overestimation of one's future consumption needs can compensate for undersaving due to hyperbolic discounting.[49] Or overconfidence might counteract lack of willpower.[50] In Besharov's illustrative model, feelings of regret – which might appear irrational because they create disutility over sunk costs – and overconfidence in one's abilities can induce someone to exert more present effort despite the existence of present-bias.[51]

Besharov's point is that intervention to correct one bias might actually reduce the individual's welfare. But set that point aside, and focus instead on the implications for future policy changes. When a new paternalist policy designed to "fix" a cognitive error is introduced, the second-best theory indicates that other problems could get worse, thus generating support for policies designed to fix them. For instance, suppose a new policy is implemented to counteract overconfidence or excessive optimism about investment opportunities. In line with Jolls and Sunstein's debiasing proposal for dangerous products, the policy might expose potential investors to horror stories about lost savings.

---

[48] Gregory Besharov, Second-Best Considerations in Correcting Cognitive Biases, 71 Southern Economic Journal 12 (2004).
[49] Matthew Rabin, Comment, in Behavioral Dimensions in Retirement Economics, Henry Aaron, ed., Brookings Institution Press and Russell Sage Foundation 247 (1999), 250-251; cited in Besharov 2004, 12-13.
[50] Roland Benabou and Jean Tirole, Self-Confidence and Personal Motivation, 117 Quarterly Journal of Economics 871 (2002); cited in Besharov 2004, 13.
[51] Besharov 2004, 15-16.

This policy might successfully reduce overconfidence, hence reducing the person's perceived benefit of saving and investing *at all*, and thereby exacerbating the undersaving problem created by hyperbolic discounting. This will tend to increase the demand for policies to counteract undersaving. And those policies might have yet other effects, as yet unforeseen, if hyperbolic discounting offsets still other cognitive biases.

Some new paternalists might actually be happy with the process described: the state's correction of one bias creates the incentive to correct other biases, until all the biases are corrected. But others, who might have been persuaded by the new paternalist's insistence on the modesty of his proposals, should be less sanguine. The second-best problem emphasizes the potential for increasing involvement of the state in cognitive correction efforts. What starts as a single targeted intervention could escalate into a far more ambitious project. There is also no reason to assume that subsequent corrective policies, whose purpose is to correct problems exacerbated by old ones, will necessarily fit the new paternalist mold. When a problem is relatively minor, decisionmakers will be inclined to support only modest intervention; when a problem looms larger, decisionmakers might support more intrusive interventions. Those who favor small interventions *considered in isolation* might reconsider that support in light of the bigger picture.

Offloading of taxes to the future. The advocates of sin taxes to correct for self-control problems assume that the affected person will respond to the taxes by reducing consumption. This conclusion does not necessarily follow when people are not perfectly rational, as they may have other self-control problems that impede their response to the tax. For instance, someone who is willing to impose health costs on her future self (by

overeating now) might also be willing to impose financial costs on her future self (by reducing her saving, or by charging the snacks to a credit card). This person could simply offload the burden of sin taxes to the future.[52]

Here again, the attempt to correct one problem could make other problems worse. The slippery slope risk emerges if the worsened problem creates demand for further intervention. In this case, a corrective sin tax could exacerbate the problem of undersaving, thereby creating support for further intervention to manipulate savings behavior. Of course, the steps in the process are not given, and the slippery slope not guaranteed. Whether the sin tax leads to reduction of consumption or offloading of the tax – or some of both – depends on the characteristics of the specific individual's bias. The tax might succeed for some and fail for others. Even if it fails, that failure will not necessarily lead to further interventions. The broader point, arising from this point and the previous point on second-best problems, is that paternalist interventions will generate unintended consequences through their effects on economic incentives. The resulting changes in behavior can lay the groundwork for further interventions.

Reduced incentives to learn. The new paternalists' leading example of successful paternalism (notably, non-governmental paternalism) is default enrollment in savings plans, which substantially increases enrollment rates.[53] But as the new paternalists also admit, default enrollments have had an unintended consequence: those automatically enrolled stick with the default asset allocation as well.[54] Because of the generally low returns to the default allocations, Choi, et al. (as cited by Camerer, et al.) found that automatic enrollment produced offsetting effects: "While higher participation

[52] Whitman 2006, 11, 12.
[53] Camerer, et al., 2003, 1227; Thaler and Sunstein 2003, 176-177.
[54] Camerer, et al., 2003, 1228, citing Choi, et al. (2002), infra.

rates promote wealth accumulation, the lower default savings rate and the conservative default investment fund undercut accumulation," and in their sample the two effects were approximately equal in magnitude.[55] Under the original policy of enrollment by active choice, those who chose actively had an incentive to pick a good allocation as well. Under the new policy, that incentive is lessened, since default enrollment in some plan reduces the costliness of failing to educate oneself about better plans.

The path to future policy changes is clear. It is not enough to implement default savings; the default allocation must be selected as well. Now, it is certainly possible to leave the allocation at the conservative, low-return level. But given the initial justification for having default enrollment at all – the desire to increase savings – further regulation follows naturally from the initial policy decision. A careful analyst will argue that the original goal was not to increase savings *per se*, but to correct a bias; once the bias is corrected, the job is finished. But here vagueness comes into play. Theoretically, in the presence of context-dependent preferences, we lack a clear standard for bias-free decisionmaking. And empirically, even if such a standard does exist, real-world decisionmakers have no means to apply it; the correct policy depends on knowledge they lack. The unchanged rate of overall wealth accumulation could easily be taken as evidence of remaining bias that requires correction (on the same grounds as the original bias).

The generalized moral hazard problem. This example illustrates a more important point: self-awareness and self-correction are skills that must be learned. People who know they will bear the consequences of their own cognitive errors have an

[55] James J. Choi, David Laibson, Brigitte Madrian, and Andrew Metrick, For Better or For Worse: Default Effects and 401(k) Savings Behavior, Pension Research Council, Working Paper No. 2002-2 (2002).

incentive to learn self-management techniques. This does not mean they always succeed, but it does mean we should expect less learning to occur in the presence of policies that reduce the cost of failure. Default enrollment reduces the incentive to learn about good investment choices. Similarly, other policies that substitute for self-correction will tend to reduce self-correction skills, which can have impacts on other aspects of personal choice. For example, if people come to expect protection against their excessive optimism, they have less reason to acquire critical thinking skills that will guard against both optimism and other errors of information processing. If people come to rely on policies that substitute for willpower, they have less reason to develop that willpower to begin with. Jonathan Klick and Gregory Mitchell refer to such effects as the "moral and cognitive hazards" of paternalistic intervention.[56] The slippery slope risk emerges because failure to learn self-management techniques can lead to more errors of judgment, which then are used to justify further interventions.

Furthermore, people's failure to learn self-control and self-correction skills can result in a "spillover" effect, as additional cognitive errors may occur not just in the area of the original policy, but in other areas as well. The reason, as Klick and Mitchell observe, is that some forms of learning are domain-general:

> For instance, developing effective self-control techniques in order to save for an automobile or home may generalize to effective strategies for retirement saving. Or, as demonstrated by empirical research on the endowment effect, people may learn to overcome consumer biases with greater market experience, and this learning may generalize across goods.[57]

If new paternalist policies decrease the need to engage in certain kinds of learning, the result could poorer performance in other, as-yet-unregulated aspects of life. This effect

[56] Jonathan Klick and Gregory Mitchell, Government Regulation of Irrationality: Moral and Cognitive Hazards, 90 Minn. L. Rev. 1620 (2006).
[57] Klick and Mitchell 2006, 1631.

might be considered a direct argument against the initial paternalist policies, but that is not our point here; we are concerned with the how implementing the initial policies increases the likelihood of implementing others. Decisionmakers who have bought the new paternalist line – that cognitive errors justify intervention – will then tend to support additional policies to deal with the newly emerging errors in choice and judgment.

B. Enforcement Need Slopes

Eugene Volokh points out the potential for slippery slopes when at least some decisionmakers view the (apparent) failure of one intervention as justification for further intervention; often, the second intervention is justified on grounds of the need to enforce the first.[58] His leading example is marijuana policy: some people might not initially support making marijuana illegal, but once it is illegal, they take the position that the law ought to be enforced rigorously (perhaps to avoid disrespect for the law).[59]

Attaining the perceived goal. New paternalism is vulnerable to enforcement need slopes because some modest initial proposals will have only modest success (or worse) at achieving their perceived goals. The problem with default savings plans leading to reliance on the default asset allocation, discussed earlier, might provide the seed of an enforcement need slope. If the initial goal is seen as increasing savings, and the overall savings rate fails to rise enough, then some decisionmakers will call for regulation of asset allocation. If that measure also fails – perhaps because people become more inclined to opt out when the contribution rate is larger – then some decisionmakers might suggest that the default plan become mandatory.

[58] Volokh 2003, 1051ff.
[59] Volokh 2003, 1051-1052.

Crowding out. Another potential source of initial policy failure is that paternalist policies could "crowd out" self-correction efforts. This is similar to the earlier point about reduced incentives to learn self-correction techniques, but the economic mechanism at work is different. The literature on public goods reveals that state funding of public goods can crowd out private funding, which means the state cannot simply fill in the gap between current funding and optimal funding – it has to provide more and more funding as the private sector provides less and less.[60] James Buchanan[61] has made a similar point about Pigovian taxes designed to internalize negative externalities such as pollution: To the extent that the polluters already care about the ill effects of their behavior (even if they care less than they should), they will have already controlled their behavior to some degree. If a tax is imposed to deal with the same problem, the polluter might decrease his self-correction because he sees the tax as performing the same job.[62]

How would new paternalist policy lead to crowding out? Presumably, even hyperbolic discounters care at least *some* about their future selves (or about their long-run interests), although perhaps less than they should. That caring is implemented via willpower and self-imposed rules. The self-imposed rules can take various forms: resolutions, limitations on refrigerator contents, and commitments to third parties (like family members or Alcoholics Anonymous).[63] Policymakers devising policies to correct for self-control problem should, presumably, take these self-correction efforts into account. The problem, however, is that the individual may respond by reducing the

[60] See Burton Abrams and 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).
[61] James Buchanan, Cost and Choice (1969).
[62] Buchanan 1969, 76-80.
[63] Whitman 2006, 7-9.

extent of their "caring" and associated self-control efforts. If the individual regards internal correction and external correction as substitutes, as some research indicates to be the case[64], the latter will tend to crowd out the former.

To the extent that crowding out occurs, the initial policy will be ineffective. The initial policy merely had to address the gap between the individual's level of self-correction and the policymaker's ideal. But if crowding out occurs, the gap will remain, thus providing a justification for yet further intervention – in the form of a higher tax or more intrusive regulation designed to force compliance.

C. The *Ad Verecundiam* Heuristic

A key insight of behavioral economics is that people's attitudes are context-dependent. Susceptibility to framing is one example; status quo bias is another. Both effects can be traced, at least in some cases, to an attempt by uninformed and boundedly rational people to glean information. When one savings plan is chosen as the default over others, for instance, employees who would prefer not to spend energy researching investment options may assume (perhaps unconsciously) that someone with expertise must have thought the default plan was a good one.

In the political and legal spheres, wherein most people are ignorant and lack strong incentives to become informed, the tendency to defer to experts can be even stronger. As one example, Volokh offers the proper scope of police warrants: regular citizens unfamiliar with the law or police tactics will be inclined to assume the experts

[64] Ayelet Fishbach and Yaacov Trope, The Substitutability of External Control and Self-Control, 41 Journal of Experimental Social Psychology 256 (2005).

(judges) have probably arrived at reasonable rules.[65] We can draw a general lesson from the example:

> We should expect attitude-altering slippery slopes to be more likely *in areas that are viewed as complex, or as calling for expert factual or moral judgment*. The more complicated a question seems, the more likely it is that voters will assume that they can't figure it out for themselves and should therefore defer to the expert judgment of authoritative institutions, such as legislatures or courts.[66]

We could also add scientists, economists, and legal scholars to the list of authorities. We will dub this tendency to defer to authorities, of whatever variety, the "ad verecundiam" heuristic (after the Latin for the "appeal to authority," a traditional fallacy of logic).

New paternalist proposals, based on the insights of these academic authorities, may make policymakers, judges, and the general public more inclined to defer to the perceived wisdom of the experts in social science and cognitive science. We should therefore ask, what ideas may become entrenched because people internalize the perceived opinions of such experts?

One idea conveyed by the new paternalism is that experts have identified *objectively correct* notions of human welfare. This is distinct from the notion of *subjective* welfare that has historically reigned in economics, where individual preferences are generally treated as given, and to a lesser extent in law, where contract law, in particular, relies on advancing the interests and expectations of the parties as they perceive them (or perceived them at the time of signing).

Now, the new paternalists may not *intend* to send this message; in some passages, they seem only to want to advance the true subjective interests of the people affected – to give them, as the Spice Girls would say, what they *really* really want. Sunstein and

[65] Volokh 2003, 1080.
[66] Volokh 2003, 1082; emphasis in original.

Thaler define "inferior decisions in terms of their own welfare" as "decisions that they would change if they had complete information, unlimited cognitive abilities, and no lack of self-control."[67] But what would they in fact choose under those conditions – what do they actually prefer? As noted earlier[68], Sunstein and Thaler also emphasize repeatedly that when decisions are context-dependent, the very meaning of individual preferences is in doubt. There seems to be internal conflict among distinct and unrationalized preference sets – and in such cases, the new paternalists do not hesitate to choose among them. Although there is no strong theoretical basis for that choice (as we argued in Part II, B and C), non-academics could hardly be blamed for thinking the choice must be justified somehow; these are the experts, after all.

In their specific policy proposals, the new paternalists regularly make judgments about which frame of reference is best by reference to the actual choice favored by it. Sunstein and Thaler rely on differences between willingness-to-pay and willingness-to-accept to explain the efficacy of changes in the default rules of contract; and then they implicitly assume that certain contractual requirements – greater vacation time, for-cause dismissal, specific safety measures, and so on – are the preferred outcomes.[69] This conclusion is by no means obvious, once we realize that other contractual terms such as the pay rate will likely adjust to account for the added benefits and guarantees.

The analytical wedge that allows the new paternalists to say people are making cognitive errors is the existence of *within-person inconsistencies* of choice, usually identified in experimental or laboratory contexts. But in their writing, the new paternalists frequently refer to objective factors about choices (without any visible

[67] Sunstein and Thaler 2003, 1162.
[68] See supra notes 27 and 28 and accompanying text.
[69] Sunstein and Thaler 2003, 1174-1177.

inconsistency) as *ipso facto* evidence of irrationality. Camerer, et al., in discussing default contributions to 401(k)'s, treat it as obvious that savings needs to be increased, based on macroeconomic concerns as well as "people's self-reports that they save less than they would like."[70] Macroeconomic concerns do not demonstrate an individual decision failure; nor do survey responses, once we recall that talk is cheap. Similarly, Thaler and Sunstein point to obesity rates as evidence of decision failure:

> "However, studies of actual choices for high stakes reveal many of the same problems [as in experiments]. For example, the Surgeon General reports that 61 percent of Americans are either overweight or obese. Given the adverse effects obesity has on health, it is hard to claim that Americans are eating optimal diets.[71]

Overweightness and obesity *per se* cannot demonstrate an inconsistency of choice; for some people, the subjective gains from heavy eating could outweigh their health concerns. It is worth noting that obesity and overweightness have both increased during the same time period that many of the associated health risks – such as heart disease – have rapidly declined.[72] In a different paper, Sunstein and Thaler cite the same health statistics, but then admit our point:

> 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.[73]

In this version of their argument they emphasize the subjectivity of the decision; yet they still rely on sheer numbers as evidence for the existence of irrationality. We consider it telling that in the earlier version they don't even make these caveats. It is easy to see

[70] Camerer, et al., 2003, 1227-1228.
[71] Thaler and Sunstein 2003, 1167.
[72] [Insert reference]
[73] Sunstein and Thaler 2003, 1168.

how statements like these will tend to be perceived as an endorsement of health as the sole appropriate measure of welfare.

The new paternalists' assumptions about what is objectively best do not appear only in their verbal statements, but in their models as well. Jonathan Gruber and Botond Koszegi, in justifying the correction of "internalities" of smoking by means of cigarette taxes, assume (without argument) that "the agent's long-run preferences [are] those relevant for social welfare maximization."[74] That assumption is crucial to the objective conclusions of their mathematical model. Ted O'Donoghue and Matthew Rabin make the same assumption in their model of "optimal sin taxes."[75]

Again, we should emphasize that theory shows only the existence of internal inconsistency, not a means of choosing one preference set over another. Nevertheless, the experts, through both their words and modeling choices, seemingly assent to the notion of objectively correct preferences or objectively desirable goods. If new paternalist policies are implemented, these assumptions will become enshrined in law. The *ad verecundiam* heuristic will apply doubly – because of the expertise of the academics, and the added authority of policymakers, judges, and bureaucrats. That, in turn, could increase support for yet more paternalist policies based on the notion that policy can and should promote objective goods and preferences, *whether or not there is any demonstrable inconsistency of individual choice*. The new policies justified by the inferred principle of objective goods need not be modest in character, as the principle in question can justify much more. The proponents of the new policies need only point to

[74] Gruber and Koszegi 2001, 1287.
[75] O'Donoghue and Rabin 2003b, 5.

previously established policies to demonstrate the acceptability of favoring supposedly objective values.

D. Preference for Simple Principles

Slippery slope analysts have often observed the tendency for subtle and complex principles to get pared down to much simpler principles. Eric Lode quotes Justice Cardozo's observation 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."[76] Frederick Schauer takes note of the "bias in favor of simple principles"[77] in law. Volokh observes a similar bias at work 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."[78]

Why do decisionmakers display this tendency? People will often look to existing policies and rules and infer the justifications directly from them. If they do look to the original debates, they will often try to summarize them quickly, drawing out what they see as the most salient details. But the process is imperfect. An original policy P1 might have been supported by a relatively narrow justification J1, while a broader justification J2 would have justified both P1 *and* P2. Looking back, the observer might incorrectly – or opportunistically – infer that J2 was the real reason for P1's enactment. The result is a broadening of the original principle.

---

[76] Lode 1999, 1516.
[77] Schauer 1985, 372.
[78] Volokh 2003, 1089.

The application to the new paternalism is straightforward. To justify their policies, the new paternalists point to the existence of internally inconsistent choices. But as we observed earlier, their presentation of the argument is not always clear; they at least appear to endorse favoring some preferences over others. After the proposals have been implemented, and more intrusive policies are on the table, what inference will be drawn from the less intrusive policies already in effect? A simplistic summary of the new paternalist argument would strip out all reference to internal conflict, and focus instead on the notion that we can justifiably choose among preferences. An even greater simplification would focus on the perceived goals of the new paternalist polices: to induce greater savings, to encourage better health choices, to support certain desirable terms in contracts, and so forth.

A variant of the bias toward simple principles is the tendency to pare multiple justifications down to one. An initial policy P1 might be supported by multiple justifications J1 and J2. A later proposal P2 might be supported only by J1. People looking back on the passage of P1 might simplify the decision by ignoring J2 and treating J1 as the sole justification.

New paternalist laws often draw additional support from the existence of other, non-paternalist arguments. For instance, laws designed to encourage healthier or less risky choices are attractive not merely because they might help the choosing individuals, but also because they reduce the burden on public health systems. Helmet laws are justified in part by paternalism (protecting the motorcycle rider from his own foolish choices) and in part by the cost helmetless riders impose on public emergency rooms.[79]

[79] See Wendy Max et al., Putting a Lid on Injury Costs: The Economic Impact of the California Motorcycle Helmet Law, 45 J. Trauma: Injury, Infection & Critical Care 550 (1998).

The prohibition is supported initially by a dual justification: "the activity imposes harm on others, and probably isn't good for the individual anyway." Later, however, the justification may be reduced to "it's okay to restrict the individual for his own good." That, of course, is a principle that can justify intervention even when the benefits to others are small or non-existent.

Purely rational, perfectly informed, and cognitively unbounded policymakers, judges, and voters would not make mistakes like these. They would evaluate each policy carefully, cogitate on the principle or principles that would justify it, consider their own independently-chosen values, and make a decision on the merits. But as the new paternalists remind us, people are not like that. Having limited information and bounded cognitive powers, they will economize by employing heuristics to decide on new policies and cases. As a result, they are likely to internalize principles embodied by the status quo – a point we made when discussing the *ad verecundiam* bias. Moreover, they will not necessarily internalize the nuanced principles of their predecessors; instead, they will often internalize stripped-down and simplistic versions of those principles. The entrenchment of less sophisticated principles lays the foundation for more intrusive and less desirable policies.

E. Framing Effects and the Shifting Status Quo

As discussed in the introduction, the new paternalists often draw attention to the moderate character of their proposals. References to the "middle ground" or "middle course" are common. A passage from Camerer, et al. (quoted more briefly in the introduction) captures the rhetorical flavor of the movement:

> 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.[80]

This form of argument exploits a cognitive bias of which the new paternalists are surely aware: the power of framing to change what is seen as moderate or extreme. Proposals are more likely to be accepted when presented in the context of more extreme positions on either side; Itamar Simonson and Amos Tversky dub this tendency "extremeness aversion."[81] Like Goldilocks choosing amongst the Three Bears' beds, people presented with soft, medium, and hard options will tend to choose medium.

This kind of framing effect can be used to indict market outcomes. For instance, in a study in which potential camera buyers were presented with two options, a low-end camera and a mid-level camera, half of the customers chose the low-end camera as the better deal; but when presented with three options, a low-end, a mid-level, *and* a high-end camera, many more customers chose the mid-level over the low-end camera.[82] Marketers could take advantage of this effect to get customers to buy more expensive products, and this is presumably the kind of behavior that new paternalists would like to change. But the very same kind of framing effect can occur in political and legal contexts. Deliberately or not, the new paternalists have framed the discussion in a way likely to make their proposals more attractive.

More importantly, in the context of slippery slopes, the *implementation* of their policies would reframe the political and legal debate. As framed by the proponents, new

[80] Camerer, et al., 2003, 1212-1213.
[81] Itamar Simonson and Amos Tversky, Choice in Context: Tradeoff Contrast and Extremeness Aversion, 29 J. Marketing Res. 281 (1992).
[82] Simonson and Tversky 1992, 290; cited in Volokh 2003, 1101.

paternalist policies lie at the "center" of the debate, between laissez-faire and more intrusive paternalism. But once passed, they would cease to be the center. Somewhat more intrusive proposals would take center stage, book-ended by existing new paternalist policies on the left and yet more intrusive proposals on the right. The new "moderate" would no longer be soft paternalism, but (let us call it) medium paternalism.

The treatment of cigarette smoking is one area in which this kind of effect has occurred. When the first cigarette bans were introduced, for airplanes and workplaces, few contemplated further restrictions. The airplane and workplace bans were the middle ground between laissez-faire and more extensive prohibition. Now, however, workplace and airplane bans are taken as given, and the focus has shifted to smoking bans in indoor restaurants and bars. Such bans are positioned as the middle ground between the extremes of "only" banning in planes and workplaces, on the one hand, and implementing wider-reaching bans on the other. And in California, where the ban in indoor restaurants and bars is status quo, some localities are now considering (and passing!) bans on smoking in outdoor locations, including restaurant patios, sidewalks, and beaches. The progression aptly demonstrates how new policies can change the status quo, so that proposals once regarded as the extreme come to be regarded as the middle ground.

The smoking example also illustrates the bias toward simple principles. Bans in workplaces and airplanes were justified primarily on the basis of non-smokers being exposed to second-hand smoke in an enclosed space, with great sacrifices needed to avoid it: don't travel by plane, work someplace else. The bans in restaurants and bars have been justified on similar grounds, but with a much less severe sacrifice: go eat or

drink somewhere else. For the beach, sidewalk, and patio bans, the sacrifice necessary to avoid second-hand smoke is the same, but the enclosed-space justification has been lost. The apparent direction of change is toward justifications that require smaller and smaller benefits to others, combined with the paternalist justification that the smokers shouldn't smoke anyway.

The general point is that the supposedly moderate character of new paternalist policies does not guarantee their staying power. The very passage of such policies reframes the political debate in way that makes further changes in the same direction more likely.

IV. Conclusion: Reasonable Expectations about Decisionmakers

The existence of a slippery slope risk does not, of course, constitute a knock-down argument against any and all new paternalist proposals. Sufficiently great benefits can justify the risks, particularly if the risks can be minimized. There exist various means of mitigating slippery slope risks, though all such means are imperfect.[83] Exploring ways in which new paternalist policies could potentially be "immunized" against the slope risk is beyond the scope of this article; we will, however, make some broad suggestions about how recognition of the slope risk should affect our thinking about paternalism.

One lesson of behavioral economics is that we ought not expect decisionmakers to perform extensive calculations, to collect all relevant information, to ignore irrelevant information, and to make reasoned decisions in all cases. This is no less true of policymakers, judges, and bureaucrats than of consumers. Indeed, it is arguably *more*

[83] Rizzo and Whitman 2003, 578-591.

true for these groups.[84] Private actors making choices for themselves, and bearing the costs and benefits of those choices, at least have the incentive to root out their errors and correct them. That does not mean they will always succeed. But at least the effects of their errors are relatively localized, and they can select courses of corrective action (also possibly in error) that take account of their personal characteristics and special circumstances. Public decisionmakers, by contrast, do not face all the costs and benefits of their choices. They make choices that create costs and benefits for numerous people besides themselves, including future generations, and they have the capacity to impose these choices society-wide. Even traditional economic theory, with its rational-actor model, does not predict wise and efficient policymaking under these circumstances.

The new paternalists have thus far paid little attention to these factors. They apparently hope policymakers will dutifully study the economic, scientific, and psychological research that identifies the existence of cognitive biases, their extent, and their locus; and then carefully craft policies designed to target those individuals in need while minimizing harm to others. That is the basic prescription of "asymmetric paternalism," for instance.[85] This ideal of new paternalist decisionmaking stands in sharp contrast to the blunt-instrument approach exemplified by recent proposals to ban trans-

[84] Edward L. Glaeser, Paternalism and Psychology, 73(1) University of Chicago L. Rev. 133 (2006), Bryan Caplan, Rational Ignorance versus Rational Irrationality, 54(1) Kyklos 3 (2001); Bryan Caplan, Rational Irrationality: A Framework for the Neoclassical-Behavioral Debate, 26(2) Eastern Economic Journal 191 (2000).
[85] Camerer, et al., 2003, 1212.

fats in Chicago and New York[86], or to ban all smoking in public places in parts of California.[87]

If we are to resist slippery slopes, then, we need to employ reasonable models of how public decisionmakers behave. That means we cannot expect them to make fine distinctions, to implement nuanced decision rules, and to engage in careful balancing of empirically verifiable needs based on valid theoretical reasoning. To expect otherwise is to ignore the central findings of both traditional economic theory and behavioral economics.

[86] Thomas J. Lueck and Kim Severson, New York Bans Most Trans Fats in Restaurants, The New York Times, December 6, 2006, downloaded January 18, 2006 from http://www.nytimes.com/2006/12/06/nyregion/06fat.html?ex=1323061200&en=aecf1ac963f8435d&ei=5088partner=rssnyt&emc=rss.

[87] Josh Gerstein, A Coastal City Bans Outdoor Smoking in Public Places, The New York Sun, March 17, 2006, downloaded January 18, 2006 from http://www.nysun.com/article/29317.