From the SelectedWorks of Mario Rizzo

October 2009

The Knowledge Problem of the New Paternalism

Contact Author

Start Your Own SelectedWorks

Notify Me of New Work

Available at: http://works.bepress.com/mario_rizzo/29

The Knowledge Problem of New Paternalism

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

It is a standing topic of complaint, that a man knows too little of himself. Be it so: but is it so certain that the legislator must know more? It is plain, that of individuals the legislator can know nothing: concerning those points of conduct which depend upon the particular circumstances of each individual, it is plain, therefore, that he can determine nothing to advantage.

Jeremy Bentham[1]

It may be admitted that, so far as scientific knowledge is concerned, a body of suitably chosen experts may be in the best position to command all the best knowledge available . . . . [Yet] scientific knowledge is not the sum of all knowledge. . . . [A] little reflection will show that there is . . . the knowledge of the particular circumstances of time and place. It is with respect to this that practically every individual has some advantage over all others in that he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active cooperation.

Friedrich A. Hayek[2]

I. INTRODUCTION 103
II. PATERNALIST WELFARE STANDARDS 106
III. PATERNALIST POLICIES ALLEGEDLY JUSTIFIED BY BEHAVIORAL ECONOMICS 107
A. Sin Taxes 108
B. Default Enrollment in Savings Plans 110
C. Cooling-Off Periods 111
D. Risk Narratives 111

* Department of Economics, New York University.

** Department of Economics, California State University, Northridge.

1. JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 319 (W. Harrison ed., Hafner Press 1948) (1781).

2. F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 521–22 (1945) [hereinafter Hayek, *The Use of Knowledge in Society*].

E. Employee-Friendly Terms in Labor Contracts 113
IV. A BRIEF THEORY OF PREFERENCE, CHOICE, AND WELFARE..... 114
V. THE PATERNALIST'S DILEMMA..... 116
VI. IDENTIFYING THE AGENT'S TRUE PREFERENCES 117
A. Local Knowledge of True Preferences..... 117
B. Conflicting Preference Sets 119
1. Hyperbolic discounting..... 119
2. Framing and endowment effects..... 123
3. Hot and cold states..... 124
VII. DISCOVERING THE EXTENT OF DECISION-MAKING BIAS..... 126
A. Lack of Precision in Measuring Extent of Bias..... 126
1. Hyperbolic discounting..... 127
2. Status quo bias 130
3. Endowment effects 131
4. Hot and cold states..... 132
5. Optimism and availability bias 134
B. Absence of a Single Measure of Bias, Even Intrapersonally 135
1. Hyperbolic discounting..... 135
2. Hot and cold states..... 136
VIII. ACCOUNTING FOR SELF-DEBIASING 137
A. The Many Varieties of Self-Debiasing and Self-Regulation..... 137
1. Cognitive strategies 138
2. Environmental strategies 139
3. Directly behavioral strategies 139
B. The Significance of Context for Self-Regulation..... 140
1. Self-regulation in the laboratory versus in the wild..... 140
2. The automaticity of unconscious self-regulation..... 143
IX. ACCOUNTING FOR INTERDEPENDENT BIASES..... 144
A. Qualitative Effects..... 145
B. Quantitative Effects..... 146
X. ANTICIPATING UNRAVELING OF SELF-REGULATION AND THE SPREAD OF BIASES 148
A. Substitution Effects Between Internal and External Debiasing 148
B. Generalized Reduction of Self-Regulation..... 151
XI. ACCOUNTING FOR HETEROGENEITY: THE ONE-SIZE-FITS-ALL PROBLEM..... 153
A. Problems of Over-Inclusion and Under-Inclusion 153

B. Heterogeneity on Multiple Dimensions..... 156
1. Fraction of individuals exhibiting a type of bias 157
2. Extent of bias 157
3. Extent of self-debiasing 157
4. Degree of responsiveness to corrective measures..... 158
5. Susceptibility of self-debiasing to unraveling..... 158
XII. CONCLUSIONS: THE ROAD BACK TO OLD PATERNALISM 158

I. INTRODUCTION

Recent work in behavioral economics has given rise to a new theoretical basis for paternalist government policies.[3] The literature of behavioral economics claims that individuals may not always act in their own best interests. People are not fully "rational," as economists understand that term, because their choices are adversely affected by various cognitive biases, insufficient willpower, and difficulties of information processing. To the extent that such decision-making problems are systematic, the claim is made that deliberate structuring of decision contexts—such as by assigning appropriate default options, providing cooling-off periods for commitments, imposing sin taxes, and so forth—can in principle enhance individuals' welfare.

The "new" paternalism purports to differ significantly from more traditional paternalism. The "old" paternalism, which often grew out of moral or religious notions of the good, effectively ignored the preferences (or interests or pleasures) of the individual in favor of the preferences of the policymaker. It does not matter if the individual really enjoys consuming alcohol, says the old paternalism, because that is simply a bad preference. The new paternalism, by contrast, takes the individual's own subjective preferences as the basis for policy

3. *See generally* Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue & Mathew Rabin, *Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism,"* 151 U. PA. L. REV. 1211 (2003); Jonathan Gruber & Botond Köszegi, *Is Addiction 'Rational'? Theory and Evidence*, 116 Q. J. ECON. 1261 (2001); Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, 35 J. LEGAL STUD. 199 (2006); Ted O'Donoghue & Matthew Rabin, *Optimal Sin Taxes*, 90 J. PUB. ECON. 1825 (2006) [hereinafter O'Donoghue & Rabin, *Optimal Sin Taxes*]; Ted O'Donoghue & Matthew Rabin, *Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes*, 93 AM. ECON. REV. 186 (2003) [hereinafter O'Donoghue & Rabin, *Studying Optimal Paternalism*]; Cass R. Sunstein & Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159 (2003) [hereinafter Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*]; Richard H. Thaler & Cass R. Sunstein, *Libertarian Paternalism*, 93 AM. ECON. REV. 175 (2003) [hereinafter Thaler & Sunstein, *Libertarian Paternalism*].

recommendations. New paternalist policies allegedly help the individual to better achieve his own subjective well-being, which cognitive impediments prevent him from attaining on his own. The individual who drinks to excess, for instance, may actually be harming himself *by his own internal standards*, which he needs help in meeting because of his lack of willpower.

Many have raised objections to the use of behavioral economics to justify paternalism. At the most abstract level, there are serious philosophical questions about the welfare standards implicit in the new paternalism.[4] For instance, is the appropriate goal to maximize the hedonic satisfaction of agents, or to maximize the satisfaction of subjective preferences that can transcend hedonic considerations?[5] In this article, however, we wish to set aside the philosophical critique—at least as much as possible—and focus on a question of application: *can policymakers reasonably be expected to implement welfare-improving paternalist policies*?

Even this question is too broad, because paternalist policymaking can be criticized in various ways. A public-choice critique of the new paternalism would ask whether policymakers have the right incentives to implement wise policies, given their own self-interest and the lobbying efforts of interested parties.[6] A comparative institutional critique would observe that policymakers also have cognitive biases that could inhibit good decision-making.[7] A dynamic critique would highlight the potential for a slippery slope from modest paternalist policies to more intrusive ones.[8] Again, we will set all these critiques aside, because our question is narrower: *do policymakers have access to the knowledge needed to*

4. *See* Gregory Mitchell, *Libertarian Paternalism Is an Oxymoron*, 99 NW. U. L. REV. 1245, 1260–69 (2005).

5. *See generally* Mario J. Rizzo & Douglas Glen Whitman, Meet the New Boss, Same as the Old Boss: A Critique of the New Paternalism (Jan. 3, 2007) (unpublished manuscript, on file with New York University).

6. *See* Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133, 144–49 (2006).

7. *See generally id.* For more informal critiques of the new paternalism (especially Sunstein & Thaler's *Libertarian Paternalism*), see the weblog posts: Posting of Gary Becker, Libertarian Paternalism: A Critique—BECKER, The Becker-Posner Blog, http://www.becker-posner-blog.com/archives/2007/01/libertarian_pat_1.html (Jan. 14, 2007, 22:07 CST); Posting of Richard Posner, Libertarian Paternalism—Posner's Comment, The Becker-Posner Blog, http://www.becker-posner-blog.com/archives/2007/01/ libertarian_pat.html (Jan. 14, 2007, 21:58 CST).

8. *See* Douglas Glen Whitman & Mario J. Rizzo, *Paternalist Slopes*, 2 N.Y.U. J.L. & LIBERTY 411, 412–13 (2007); Mario J. Rizzo & Douglas Glen Whitman, *Little Brother is Watching You: New Paternalism on the Slippery Slopes*, 51 ARIZ. L. REV. (forthcoming 2009).

implement welfare-improving paternalist policies? The answer, we argue, is no.

The title of this article is inspired by the "knowledge problem" identified by Friedrich A. Hayek in his critique of socialist central planning.[9] In the early twentieth century, the advocates of socialism argued that, in principle, a central planner—equipped with all relevant knowledge of resource endowments, technologies, and preferences—could design an efficient economic plan for society.[10] In response, Hayek said that to assume the central planner possesses all the relevant information about endowments, technologies, and preferences is to *assume the problem away*.[11] The critical problem that any economic system must solve is to mobilize and use knowledge that "never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess."[12] A signal virtue of a voluntary market order is that it creates conditions under which such information is more likely to be mobilized and used.

The most important reason that many economists had failed to appreciate this knowledge problem is that they had been deceived by their own excessively simple models. They had taken models useful in understanding some limited features of the real world, such as the equilibrium reaction of markets to supply or demand shocks, and applied them to the broader problem of substituting government planning for market processes. They were guilty of the fallacy of the misplaced concrete: simple models were mistaken for a simple world.[13]

We argue that the new paternalism spawned by behavioral economics faces a very similar knowledge problem and for similar reasons. *If* well-meaning policymakers possess all the relevant information about individuals' true preferences, their cognitive biases,

9. F. A. Hayek, *Socialist Calculation I: The Nature and History of the Problem* (1935), *in* COLLECTIVIST ECONOMIC PLANNING (F.A. Hayek, ed. 1935), *reprinted in* F. A. HAYEK, INDIVIDUALISM AND ECONOMIC ORDER 119 (1948); F. A. Hayek, *Socialist Calculation II: The State of the Debate* (1935), *reprinted in* HAYEK, INDIVIDUALISM AND ECONOMIC ORDER, *supra*, at 148; F. A. Hayek, *Socialist Calculation III: The Competitive "Solution"* (1940), *reprinted in* HAYEK, INDIVIDUALISM AND ECONOMIC ORDER, *supra*, at 181.

10. *See* Oskar Lange, *On the Economic Theory of Socialism: Part One*, 4 REV. ECON. STUD. 53, 68–71 (1936); Oskar Lange, *On the Economic Theory of Socialism: Part Two*, 4 REV. ECON. STUD. 123 (1937); A.P. Lerner, *A Note on Socialist Economics*, 4 REV. ECON. STUD. 72 (1936). *See generally* Hayek, *The Use of Knowledge in Society*, *supra* note 2.

11. *See* Hayek, *The Use of Knowledge in Society*, *supra* note 2, at 519.

12. *Id.*

13. *See generally* F. A. HAYEK, THE COUNTER-REVOLUTION OF SCIENCE (1952).

and the choice contexts in which they manifest themselves, *then* policymakers could potentially implement paternalist policies that improve the welfare of individuals by their own standards. But lacking such information, we cannot conclude that actual paternalism will make their decisions better; under a wide range of circumstances, it will even make them worse. New paternalists have not taken the knowledge problems that are evident from the underlying behavioral and economic research seriously enough.

To begin, we focus the discussion by outlining several specific policies that authors in the new paternalist literature have advocated, as well as the welfare standards and cognitive biases that allegedly justify them. These policies are chosen as illustrative of new paternalist policies more generally. The remainder of the article uncovers a series of knowledge-based obstacles that paternalist policies must overcome in order to be effective and justified.

Specifically, paternalist policymakers must (1) identify agents' "true" preferences that are to be maximally satisfied; (2) determine the extent of each cognitive bias or decision-making problem; (3) properly account for privately adopted self-debiasing measures, as well as how paternalist policies would affect such measures; (4) deal with the problem of interdependent biases; (5) anticipate unraveling and unlearning effects; and (6) account for heterogeneity in the population with respect to all of these factors. We argue that these factors taken together present a formidable barrier that robs the new paternalism of any presumption of welfare improvement—even if the underlying theory and empirical results of behavioral economics are granted. Furthermore, paternalist policymakers who lack the information needed to implement policies that actually assist individuals according to their own subjective preferences will tend to substitute their own.

II. PATERNALIST WELFARE STANDARDS

The "new paternalist" literature, as we shall call it, emphasizes the possibility of making individuals "better off" according to their own preferences. Richard Thaler and Cass Sunstein, for instance, adopt a welfare standard defined in terms of what people would do if they were perfectly rational:

> We intend "better off" to be measured as objectively as possible, and we clearly do not always equate revealed preference with welfare. That is, we emphasize the possibility that in some cases individuals make

> inferior choices, *choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower*.[14]

Similarly, Colin Camerer and coauthors choose as their welfare standard the decisions that individuals *would* make if they were fully rational, defined as follows:

> First, people have well-defined preferences (or goals) and make decisions to maximize those preferences. Second, those preferences accurately reflect (to the best of the person's knowledge) the true costs and benefits of the available options.[15] Third, in situations that involve uncertainty, people have well-formed beliefs about how uncertainty will resolve itself, and when new information becomes available, they update their beliefs using Bayes's law—the presumed ability to update probabilistic assessments in light of new information.[16]

The essential problem, as the new paternalists see it, is that individuals are unlikely to pursue choices that are "in their best interest"[17] in many cases because of cognitive or behavioral biases. These include "self-control problems," "fail[ure] to process information as Bayes's rule would require," and "systematic mispredictions about the costs and benefits of choices."[18]

III. PATERNALIST POLICIES ALLEGEDLY JUSTIFIED BY BEHAVIORAL ECONOMICS

A wide variety of paternalistic policies could potentially be justified using these welfare standards, but we cannot address all of them. We will therefore rely, when necessary, on five illustrative proposals: sin taxes, default enrollment in savings plans, cooling-off periods for consumer purchases, risk narratives to accompany risky products, and employee-friendly terms in labor contracts. With each proposal, we also discuss the

14. Thaler & Sunstein, *Libertarian Paternalism*, *supra* note 3, at 175 (emphasis added).

15. The second criterion seems to suggest that the agent can have less than complete knowledge so long as he makes efficient use of his incomplete knowledge. This means that true preferences are simply optimally-informed preferences. Therefore for true preferences, in this attenuated sense, to be different from actual preferences requires that the real-world individual have less than socially-optimal incentives to acquire information. The authors do not expand on this point. To use this criterion as a standard for policy intervention would require the preference paternalist to stop short of complete information in determining true preferences. How far short would be difficult to assess both theoretically and empirically.

16. Camerer et al., *supra* note 3, at 1214–15 (2003).

17. Thaler & Sunstein, *Libertarian Paternalism*, *supra* note 3, at 175. *See also* Camerer et al., *supra* note 3, at 1212.

18. Camerer et al., *supra* note 3, at 1217–18.

decision-making problems identified by behavioral economics that are used to justify the policy proposals. The problems with these diverse policies will not be identical, but all proposals will encounter at least some of the impediments outlined in the remainder of this article.

A. Sin Taxes

Some analysts, notably O'Donoghue and Rabin[19] and Gruber and Köszegi,[20] propose to impose sin taxes—e.g., a tax on fatty foods—to induce better behavior.

The behavioral justification for these sin taxes is that individuals are afflicted by present-bias or insufficient willpower. Very simply put, individuals place too much weight on the present relative to the future.[21] 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, they procrastinate, they become addicted to drugs, and so on.[22]

This is, however, a simplified account. In traditional economic theory, there is nothing *per se* irrational about placing more weight on the present than the future. Indeed, economic models of intertemporal choice almost universally assume the individual has some *discount factor* (often symbolized with the Greek letter δ) that he applies to future costs and benefits. For instance, someone with a discount factor of $\delta = 0.90$ would consider $100 of benefits to be received in a year to be equivalent to $90 received immediately. The individual's discount factor is generally considered a matter of subjective preference.[23]

For the individual's behavior to be internally consistent, however, the discount factor must be constant. That is, the trade-off between benefits at time 1 and at time 2 should depend only on their distance

19. *See* O'Donoghue & Rabin, *Optimal Sin Taxes*, *supra* note 3; O'Donoghue & Rabin, *Studying Optimal Paternalism*, *supra* note 3.

20. Gruber & Köszegi, *supra* note 3.

21. For some reason the problem of placing too much weight on the future relative to the present (hyperopia) is ignored. *See, e.g.*, Ran Kivetz & Anat Keinan, *Repenting Hyperopia: An Analysis of Self-Control Regrets*, 33 J. CONSUMER RES. 273, 282 (2006) (concluding that "consumers sometimes suffer from excessive farsightedness" and "repent hyperopia in the long run").

22. Shane Frederick, George Loewenstein & Ted O'Donoghue, *Time Discounting and Time Preference: A Critical Review*, 40 J. ECON. LITERATURE 351, 393–94 (2002).

23. Economists sometimes refer to a discount rate instead of a discount factor. The discount rate is related to the discount factor in the following way: $r = (1 - \delta)/\delta$. Throughout this paper, we will use only discount factors.

from each other, *not* on their distance from the present. Thus, for a person with a discount factor of 0.90, $100 to be received in two years should be equivalent to $90 in one year, *and* $100 to be received in one year should be equivalent to $90 now. This is known as *exponential discounting*.[24] Behavioral research, however, indicates that real people are inconsistent discounters. For instance, an individual might regard $100 to be received in two years as equivalent to $90 to be received in one, and yet he might regard $100 to be received in a year as worth only $80 now. This phenomenon is known as *hyperbolic discounting*.[25]

People who engage in hyperbolic discounting may exhibit *time inconsistency*: they will make decisions about future trade-offs and then reverse those decisions later. For instance, if offered a choice between $100 in two years and $85 in one year, the person described above chooses the larger sum. Yet when a year has passed, he reverses his prior choice and takes the smaller sum ($85), because the $100 to be received in a year is regarded as worth only $80 now. Behavioral economists take this sort of inconsistency as evidence of irrationality.[26]

Thus, proponents of sin taxes use hyperbolic discounting to explain self-control problems. Intuitively, people's inconsistent behavior reflects their vulnerability to temptation when those temptations are near. With regard to eating, for example, a hyperbolic discounter might promise to start a diet tomorrow, but then reverse that decision once tomorrow has become today. A properly-calibrated sin tax, its new paternalist supporters argue, would make the overeater fully account for the future costs of her current choices by increasing the present cost. This increase in present cost, new paternalists argue, offsets the hyperbolic discount and aligns the person's decision with "rationality."

24. "Exponential" refers to the fact that the discount factor must be multiplied by itself multiple times to discount events multiple periods in the future. For instance, in the example given, $100 to be received in two years would be valued at $(0.90)(0.90)($100) = (0.90)^2($100) = 81 now.

25. *See generally* Frederick, Loewenstein & O'Donoghue, *supra*, note 22 (reviewing the relevant literature on experiments of this nature). The seminal article in this literature is R.H. Strotz, *Myopia and Inconsistency in Dynamic Utility Maximization*, 23 REV. ECON. STUD. 165 (1955–1956); *see also* GEORGE AINSLIE, BREAKDOWN OF WILL (2001).

26. The dollar values are used here only for illustrative purposes. In principle, the costs and benefits need not be monetary; they can be pleasures, pains, health effects, and so on. The key question is how benefits and costs of whatever form are weighed against each other when they occur at different points in time.

B. Default Enrollment in Savings Plans

Various authors, but most notably Thaler and Sunstein, have advocated automatically enrolling new employees in savings plans from which they could voluntarily opt out (as opposed to the more common practice of not enrolling employees until they opt in).[27] It is not always clear in the literature whether this recommendation is directed solely at employers, or if the new paternalists would also support a government requirement that employers implement automatic enrollment. Sunstein and Thaler say the law "might require employers to provide automatic enrollment and allow employees to opt out." Further, they say this would be consistent with their notion of "libertarian paternalism," but they do not *explicitly* advocate this policy.[28] Camerer et al., strongly imply that mandatory savings default rules may be necessary because firms lack sufficient incentive to offer optimal defaults.[29] For this article, we will consider a legal mandate on employers to adopt default enrollment.

One behavioral argument in favor of default enrollment is the same as that used for sin taxes: individuals are afflicted by hyperbolic discounting, which causes them to weigh the present too heavily. The present benefits of greater consumption, combined with the present costs of going through the enrollment process, induce individuals to delay enrollment and thus save too little.

More often, however, the case for default enrollment is based on inertia or status quo bias: the psychological tendency of people to maintain current arrangements, whatever they might be.[30] Sunstein and Thaler say that employees often fail to enroll under an opt-in system, but they would choose to enroll if they simply took the time to think carefully.[31] The idea, then, is to place employees into a new status quo that is more likely to match their considered preferences.[32]

27. *See* Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1159–1202; Thaler & Sunstein, *Libertarian Paternalism*, *supra* note 3, at 175–79.

28. Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1176–77.

29. Camerer et al., *supra* note 3, at 1251–52.

30. For a detailed discussion see William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 33–41 (1988).

31. "If employers think (correctly, we believe) that most employees would prefer to join the 401(k) plan if they took the time to think about it and did not lose the enrollment form, then by choosing automatic enrollment, they are acting paternalistically by our definition of the term." Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1172–73.

32. Status quo bias and hyperbolic discounting are not always clearly distinguished in the case for default savings enrollment. For instance, Sunstein and Thaler state that "[e]ven a trivial

C. Cooling-Off Periods

There are two types of cooling-off periods. One kind creates a mandatory waiting period before a purchase or other decision can be made.[33] The other creates a mandatory period following a purchase or other decision during which it can be reversed by one of the parties.[34] For example, a cooling-off period for marriage requires a certain number of days to pass between issuance of a marriage license and the marriage itself; a cooling-off period for new cars allows a car buyer to return the car within a few days of the sale without penalty.[35]

The behavioral support for cooling-off periods is the evidence that people make different decisions depending on whether they are in a "hot" or "cool" state.[36] According to Sunstein and Thaler, "[t]he essential rationale [for cooling-off periods] is that under the heat of the moment, consumers might make ill-considered or improvident decisions."[37] Camerer et al., note that this rationale is supported by evidence that people make costly or even irreversible choices when they are in a biologically "hot" state (such as anger, fear, excitement, or sexual arousal) that they would not make if they were in a "cool state" (calm, reflective, and sober).[38] New paternalists support the cooling-off period because it either forces the decision maker to delay his decision until he is in a cooler mental state, or allows him to reconsider once he is in such a state, thus allowing him to pursue his "true" preferences.

D. Risk Narratives

New paternalists also support the use of "risk narratives" to aid individuals in making risky decisions. When consumers consider purchasing dangerous products or engaging in dangerous activities, they could be informed about the relevant risks by means of statistical

action, such as filling in some form and returning it, can leave room for failures due to memory lapses, sloth, and procrastination." *Id.* at 1181. Although they do not specifically invoke the notion of hyperbolic discounting, that is the leading explanation among behavioral economists for procrastination in areas such as dieting and saving.

33. Camerer et al., *supra* note 3, at 1240.

34. *Id.*

35. Without an ex post penalty, that is. The initial purchase price might be higher to account for costs associated with having a cooling-off period.

36. *See* George Loewenstein, *Emotions in Economic Theory and Economic Behavior*, 90 AM. ECON. REV. 426 (2000) [hereinafter Loewenstein, *Emotions in Economic Theory*].

37. Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1188.

38. Camerer et al., *supra* note 3, at 1238–40.

summaries of the likelihood of various harms. Alternatively, they could be informed by means of accounts, or narratives, about specific people who have suffered harm from the product or activity in question. Sunstein and Jolls propose that providers be required by law to provide such narratives:

> 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 . . . .[39]

We will refer to this policy as "risk narratives." The behavioral justification for this policy is that people are afflicted by optimism bias, which causes them to underestimate their *personal* likelihood of suffering adverse consequences.[40] To take just one example, they underestimate their chances of getting into an automobile accident.[41] As a result, they will be too likely to expose themselves to risks.

Interestingly, Jolls and Sunstein propose to correct optimism bias by exploiting a different bias: the availability heuristic,[42] which refers to the tendency to judge the probability of an event based on "an assessment of how easily examples of the event can be called to mind."[43] For instance, a person whose grandmother died from a rare, but not genetic, illness might overestimate the likelihood of contracting that illness—simply because it happened to someone he knows. Thus, risk narratives harness the availability heuristic to counter optimism bias—a person's overestimation of risk upon hearing a general warning is offset by an underestimation of risk caused by exposure to vivid stories about harmed individuals. The result, claim the new paternalists, should be an accurate risk calculation and an appropriate decision.

39. Jolls & Sunstein, *supra* note 3, at 212.

40. W. KIP VISCUSI & WESLEY A. MAGAT, LEARNING ABOUT RISK: CONSUMER AND WORKER RESPONSE TO HAZARD INFORMATION 95–96 (1987); Christine Jolls, *Behavioral Economic Analysis of Redistributive Legal Rules*, 51 VAND. L. REV. 1653, 1659–62 (1998).

41. Jolls & Sunstein, *supra* note 3, at 205.

42. *See generally* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974); Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCHOL. 207–32 (1973).

43. Jolls & Sunstein, *supra* note 3, at 204.

E. Employee-Friendly Terms in Labor Contracts

Sunstein and Thaler suggest various terms that could be included in labor contracts for the benefit of employees. For instance, they suggest making "for cause" rather than "at will" the default termination rule,[44] lengthening the presumed amount of paid vacation time[45] and presuming protection against age discrimination unless the employee waives such protection.[46]

In addition to these suggestions, in which the defaults are fully waivable, Sunstein and Thaler suggest other policies (including some existing policies) that are only partially waivable. For instance, they support the provision of the Model Employment Termination Act, which replaces "at will" with "for cause" termination.[47] This right can be waived by agreement—but only if the employer agrees to provide a severance payment (which the Model Act sets as "one month's salary for every year of employment") in the event of a not-for-cause termination.[48] Note that the Model Act does not allow employees to waive their right to "for cause" termination by negotiating for higher regular salary or for any severance pay less than one month's salary per year of employment. Similarly, Sunstein and Thaler reinforce the case for the Fair Labor Standards Act, which says that employees may not be required to work beyond 40 hours per week.[49] This provision may be waived in return for time-and-a-half pay.[50] Note that it cannot be waived for any lower rate of pay (including the regular rate), even if the employer and employee agree upon it.

The behavioral justification for changing default rules—and sometimes making the defaults costly to change—is that people are subject to *framing effects*.[51] This means their decisions tend to be sensitive to seemingly irrelevant aspects of how the choice situation is

44. Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1175.

45. *Id.* at 1176.

46. *Id.* at 1177.

47. *Id.* at 1187. Sunstein and Thaler do qualify their endorsement to some extent by admitting that provisions with substantive limitations on waiver are "less libertarian than [they] might be." *Id.*

48. *Id.* at 1187 (citing MODEL EMPLOYMENT TERMINATION ACT §§ 3(a), 4(c), *reprinted in* MARK A. ROTHSTEIN & LANCE LIEBMAN, EMPLOYMENT LAW: CASES AND MATERIALS 211 (Statutory Supp. 2003)).

49. *Id.*

50. *Id.* (citing 29 U.S.C. § 207(f) (2000)).

51. *See generally* Cleotilde Gonzalez et al., *The Framing Effect and Risky Decisions: Examining Cognitive Functions with fMRI*, 26 J. ECON. PSYCHOL. 1 (2005), *available at* http://repository.cmu.edu/cgi/viewcontent.cgi?article=1012&context=sds.

described. Probably the best known type of framing effect is the endowment effect, which refers to people's tendency to demand more compensation to give something up (their willingness to accept, or "WTA") than they would have paid to acquire that same thing (their willingness to pay, or "WTP").[52]

For rational agents, the default should not make a difference for choices (at least if transaction costs are low). But for less rational agents, the default rule can matter. Employees, for example, might demand more compensation to eliminate a "for cause" termination clause than they would sacrifice to insert it. They might demand more compensation to give up additional vacation time than they would sacrifice to acquire it. The idea, then, is to structure defaults in labor contracts to increase the likelihood of employees getting favorable terms.

IV. A BRIEF THEORY OF PREFERENCE, CHOICE, AND WELFARE

In order to convey the underlying unity of our knowledge-based critique of new paternalist policy prescriptions, it is useful to outline our views on the evidential character of individual choice. As we have seen, the core of the behavioral critique of standard welfare theory is the claim that individuals do not always reveal their true subjective preferences through their actual choices (because of inadequate willpower or knowledge or both). We do not deny this claim about actual choice.

However, the possibility that an individual's particular choices can be erroneous in this sense does not mean that we must abandon actual choice as the ultimate evidence of welfare. It simply means that we must be more inclusive about which choices are relevant. Individuals may be aware of their own lack of self-control or that they make systematic mistakes. When that is the case, we would expect them to make choices that manifest this awareness. For instance, individuals may bind themselves ex ante or acquire better information about the consequences of their actions. This perspective requires us to observe a *complex* of choices rather than a single choice. In order to make sense of the behavior at issue in this way, we must maintain the hypothesis that *some* of these related choices do, in fact, express the related preferences of

52. *See* Jolls & Sunstein, *supra* note 3, at 205. For instance, students "endowed" with a university mug demanded more to part with the mug than they would have paid to buy it. *Id.* Given the mug's low value relative to the students' wealth, the two situations are effectively identical: they are being asked to choose between a mug and money. Regardless of whether they were given the mug to begin with, both mug and money were options. Yet the students' choices differed.

individuals.[53] For example, suppose a worker is paid in cash. On his way home, he is tempted to stop by a bar and drink a good part of his salary. Since he often succumbs to this temptation, he may eventually take steps to avoid it. He may choose to take a different route home, bypassing the bar. He may request that his salary not be paid in cash or that his employer directly deposit his salary in a bank account. These choices, in conjunction with the choice-at-issue, reveal that he has a self-control problem.[54] But we cannot infer simply from the choice to drink away a good portion of his salary that his choice displays a lack of willpower or is otherwise "irrational."

The choices that reveal he has a problem, however, are simultaneously the choices that attempt to solve the problem. Does our approach, therefore, imply that the individual's solution is always sufficient? Can someone else know if there is a better way? These are separate questions. Obviously, the individual's solution may not be sufficient. There may be unexploited opportunities for gain; better techniques of self-control may exist. But how would anyone else know? For another to know, he must know the *benefit* to the individual of controlling his behavior as well as the *cost* of self-control mechanisms. Since the other person does not have access to the contents of the individual's mind, the only way to know is to offer a wider array of techniques to the individual and observe if he *chooses* one of them. But absent an actual choice by the individual, we cannot know.

Some have argued that since statements are speech-acts, we can use what people say about their choices as evidence of the possible irrationality of those choices.[55] For instance, if an individual says he wants to lose weight, this would constitute evidence of his actual preferring to do so—and thus of his "irrationality" in continuing to overeat. While we should not exclude this possibility entirely, it is important to point out that speech-acts and other choices generally have very different cost-benefit structures. The incentives to *say* something are not the same as the incentives to *do* the thing spoken of.

53. A maintained hypothesis is simply held for the moment. It may be questioned under different circumstances or at different times. Thus it does not amount to the view that any particular choice is privileged insofar as it *necessarily* reveals the subjective preferences of the individual.

54. Just as the worker has various means to deal with his self-control problem, acquisition of more information, solicitation of expert advice, and attempts to improve one's computational skills plausibly reflect the preference to improve the knowledge content of decisions.

55. *See* Andrew Caplin, *Economic Theory and Psychological Theory: Bridging the Divide*, *in* THE FOUNDATIONS OF POSITIVE AND NORMATIVE ECONOMICS: A HANDBOOK 336, 359–60 (Andrew Caplin & Andrew Schotter eds., 2008).

The individual can say, "I want to save more, but I am too weak-willed." What does this mean for purposes of economic analysis or public policy? It is entirely unclear. Is the individual expressing a preference or a simple desire? A preference reflects the willingness to incur the opportunity cost, whereas a desire is just a generally favorable attitude toward something irrespective of opportunity cost. The statement itself does not reveal a serious willingness to incur the opportunity cost of more savings. It is evidence simply of his willingness to incur the costs of the *statement* to attain *its* benefits.[56] The saying and the doing are different actions. Saying is not by itself evidence of true and comprehensive underlying preferences.

V. THE PATERNALIST'S DILEMMA

To understand the fundamental problem facing the paternalist, recall that the rationale for new paternalism is that the individual has cognitive and behavioral limitations that prevent him from either recognizing difficulties in the pursuit of his welfare or efficiently overcoming them. This implies a complex interrelationship between the knowledge needed by the paternalist and the knowledge possessed, or capable of being acquired, by the individual.

The paternalist must be smarter than the target agents. He must know their preferences better than they do in order to know just what their difficulties are and how they may be efficiently overcome. Yet both the problems and the solutions are contextual. They depend on local and personal knowledge. Thus, even if the paternalist has better theoretical knowledge about cognitive and behavioral biases, it will be of little use unless he has considerable local knowledge about a specific individual's preferences, self-control problems, available options, and so forth. Ultimately, if the goal is superior *action* on the part of agents, the superior theoretical knowledge of the paternalist cannot be directly relevant to the individual, except by way of advice. The best course of action for the individual to take will depend on what Hayek called "knowledge of the particular circumstances of time and place."[57] This includes knowledge of locally and temporally contingent external facts, facts about the individual's personal traits and, more specifically, facts

56. Individuals may wish to signal to others or to themselves that they are prudent without being prudent. Thus they are willing to incur the costs of deception for its benefits. This does not imply that they are willing to incur the costs of actual prudence for *its* benefits.

57. Hayek, *The Use of Knowledge in Society*, *supra* note 2, at 521.

about particular temptations and strategies to avoid them. Sometimes these facts are consciously held and utilized, while in other cases they may be tacitly or unconsciously held and utilized. This knowledge is largely inaccessible by paternalists and yet, without it, they cannot use their putatively superior theoretical knowledge to develop welfare-improving policies.

In the remaining sections of this paper, we will discuss in greater detail the many kinds of knowledge that paternalist policymakers would need to have in order to improve individual choices.

VI. IDENTIFYING THE AGENT'S TRUE PREFERENCES

The issue is whether policymakers—including voters, politicians, judges, and bureaucrats—can generate general, clearly articulable rules in the form of taxes, subsidies, cooling-off periods, and so forth, to counteract what would otherwise be the inferior decisions of agents. One prerequisite for welfare-improving policies is that policymakers must possess superior knowledge of people's "true" preferences—that is, the preferences the new paternalists allegedly wish to advance.

Evidence suggests that agents may not have "true" preferences at all.[58] This, in itself, presents a problem for the new paternalist paradigm; we cannot claim to be making people better according to their preferences if such preferences do not exist. But we will assume, *arguendo*, that true preferences do in fact exist. Let us first address the general question: Does the paternalist know true preferences better than the agent himself?

A. Local Knowledge of True Preferences

The relevant question is whether policymakers can be expected to have better knowledge of true preferences than the agents in question. Since "better" is defined in terms of the individual's subjective welfare (as opposed to old-style paternalism), we must compare the relative ability of individuals to make welfare-enhancing decisions for themselves with the ability of outsiders to decide on their behalf.

58. The new paternalists admit this. *See, e.g.*, Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1164 ("If the arrangement of the alternatives has a significant effect on the selections the customers make, then their true 'preferences' do not formally exist."). In general, the question of whether preferences formally exist will arise whenever individuals exhibit preference reversals or make frame-dependent choices.

Lacking direct evidence on the relative ability of individuals versus government to discover preferences, let us consider the relative ability of individuals versus close friends and family members. In a recent study, Joel Waldfogel compares the valuations by *consumers* of items they purchased themselves to their valuations of similar items purchased for them as gifts.[59] The gift-givers were friends and extended and immediate family members—in general, individuals who would likely have some personal, local, and sometimes tacit knowledge of the recipients' preferences.[60] The consumer goods in question were familiar to both giver and receiver, did not involve intertemporal choices, and were not uncertain in the sense that people did not think they were buying one thing and got another.[61] In other words, these were relatively "simple" consumption choices. If the consumers were no better or worse at determining what satisfied their preferences than the gift-givers, we should expect the consumers' ex post evaluations of the self-purchased items versus gifts to have been about equal, on average. Instead, "consumers' own purchases generate[d] between 10% and 18% more value, per dollar spent, than items received as gifts."[62]

Is it likely that the ignorance of consumers about their own preferences made them incapable of accurately evaluating the relative values of self-purchases and gifts? Perhaps they simply reaffirmed their decisions in the survey that was undertaken by Waldfogel shortly after the decisions were made. However, most of the products purchased by either party were of the type that would likely show their "true value" rather quickly—sweaters, shirts, books, CDs, jackets, hats, and so forth.[63] So it seems reasonable to expect that consumers' ignorance about their own preferences was largely resolved ex post. If this is so, then it makes sense to use consumer valuations of the relative efficiency of own-purchases and those of gift-givers. In view of the evidence that "consumers fare better [at identifying their own preferences] than all types of givers except significant others and possibly grandparents . . . it seems unlikely that an alternative chooser would do better than friends,

59. Joel Waldfogel, *Does Consumer Irrationality Trump Consumer Sovereignty?*, 87 REV. OF ECON. AND STAT. 691 (2005).

60. *Id.*

61. *Id.*

62. *Id.*

63. Most of the purchases were of this sort, that is, short-run experience goods. However, the value of a few purchases could not be immediately ascertained, such as electronics, kitchen appliances, or perhaps video games. *See id.* at 695 tbl.3.

siblings, and parents, all of whom have substantial amounts of information about the ultimate consumer's preferences."[64]

In short, even friends and family have a difficult time doing any better than the individual himself in making welfare-enhancing choices. Yet friends and family are more likely than policymakers to have the local knowledge necessary to make wise decisions. Thus, Waldfogel's study provides at least suggestive evidence of the difficulty new paternalists will face in crafting wise policies. The basic problem is that paternalist policymakers need a baseline of "true" preferences to satisfy, but the knowledge of such preferences is very hard to access. That individuals sometimes have difficulty determining their own preferences does not mean outsiders will do any better; they can also do worse.

B. Conflicting Preference Sets

We now turn to the more technical question of how policymakers might go about determining what true or informed preferences are, assuming once again that they do exist. The case for a decision-making bias is typically based on the existence of an inconsistency in individual choices, which presumably corresponds to an internal inconsistency of preferences. But identifying an inconsistency in someone's behavioral preferences (meaning those that actually determine choice) is not the same as identifying someone's true preferences. To do that, we would have to know which of the inconsistent behavioral preferences better represents the agent's actual welfare. We will consider three types of bias for which this problem arises: hyperbolic discounting, framing and endowment effects, and hot and cold state effects.

1. Hyperbolic discounting

Sometimes individuals make different choices about present versus future consumption depending on the time at which the decision is made, even if the two periods being compared do not change. To take the example given earlier, in the discussion of sin taxes, an individual today might choose $100 to be received in two years over $85 to be received in one year, and yet reverse that decision when a year has passed and choose $85 immediately over $100 the next year (even though nothing else has changed). This inconsistency in choice is modeled as an

64. *Id.* at 695.

underlying inconsistency in preferences.[65] We could assume—as do Gruber and Köszegi[66] and O'Donoghue and Rabin[67]—that the *true* preferences are those represented by the more far-sighted choice, and the question becomes how to make the near-term choices correspond to the far-sighted preferences. But what basis is there for this assumption? We could just as easily designate the more near-sighted preferences as the correct ones, and then aim to make far-term choices better correspond to them.

To put it another way, an internally consistent person would have a single discount factor. In our example, we have an individual with two discount factors: 0.90 between any two adjacent years in the future, and 0.80 between the present year and the next year. This person exhibits time inconsistency by choosing $100 over $85 when both are in the future, then reversing that decision by choosing $85 over $100 when the $85 is to be received immediately. One way to make this person internally consistent would be to make him use a discount factor of 0.90 for *all* his intertemporal decisions. Thus, he must choose the $100 later over the $85 earlier every time. But another way to make this person internally consistent would be to make him use a discount factor of 0.80 for all intertemporal decisions, so that he will choose the $85 earlier over the $100 later every time. As either of these "corrections" would make the agent's behavior consistent, we lack a means of saying which discount factor corresponds to the agent's "true" preferences, even if we concede that one of them must.

To make the problem more vexing, the paternalist may not face a choice between just two discount factors. The paternalist policymaker might favor the more far-sighted (larger) discount factor, the more near-sighted (smaller) discount factor, or *some discount factor that lies somewhere in between*, reflecting an intermediate degree of patience. If there is inconsistency between two different preference sets, there is no reason for the paternalist to assume the agent's "true" preferences must be one of those two.

Moreover, the research on time discounting does not reveal a simple binary choice process, wherein the individual applies one discount factor

65. Inconsistency in preferences, however, need not produce inconsistency in or reversal of choices. *See* HOWARD RACHLIN, THE SCIENCE OF SELF-CONTROL 39 (2000). In this case we would have "myopia," that is, a large short-run discount factor and a small long-run discount factor, without preference reversal.

66. Gruber & Köszegi, *supra* note 3.

67. O'Donoghue & Rabin, *Studying Optimal Paternalism*, *supra* note 3.

when comparing two future periods and another discount factor when comparing the present period and a future period. Instead, the discount factor varies continuously depending on how far away the nearer period is.[68] For example, the individual might apply the discount factor 0.98 when comparing rewards to be received ten versus eleven years in the future; the discount factor 0.90 when comparing rewards five versus six years in the future; the discount factor 0.80 when comparing rewards two versus three years in the future; and the discount factor 0.70 when comparing rewards now and a year from now. Thus, in economic terminology, preferences are truly *hyperbolic*, not just *quasi-hyperbolic*.[69] If finding the agent's true preferences means finding a single time-discounting factor that can be used as the basis for exponential discounting (which is time consistent), then the existence of true hyperbolic discounting means the paternalist has infinitely many different options to choose from—and no objective means of doing so.

There have, however, been some attempts to justify using the lower or long-term rate. The first is based on the assumption that individuals have "stable lifetime preferences" and thus any temporary deviation from them is a mistake.[70] This is reinforced by the idea that, for most of the future periods about which plans are made, a higher (more patient) discount factor is applied. Unfortunately, the idea of stable lifetime preferences is merely an assumption. Furthermore, to draw normative significance from the stylized fact that a higher discount factor applies to more periods than does the lower discount factor seems little more than an attempt to derive an "ought" from an "is."

Moreover, the "mistake" interpretation founders on the shoals of truly hyperbolic discounting. When there are only two discount factors, it

68. In his early work, Richard Thaler finds three effective annual discount rates ranging from 345% over a one-month horizon to 120% over a one-year horizon to only 19% over a ten-year horizon. Richard H. Thaler, *Some Empirical Evidence on Dynamic Inconsistency*, 8 ECON. LETTERS 201, 201–07 (1981).

69. *See* AINSLIE, *supra* note 25, at 28–35. "Quasi-hyperbolic" refers to a discounting process involving only two discount factors: one that applies between any two future periods (lower), and an additional discount that applies between the present and any future period (higher). "Hyperbolic" refers to a *continuously* declining discount rate as the future periods of comparison become more distant. *See* David Laibson, *Golden Eggs and Hyperbolic Discounting*, 112 Q.J. ECON. 443, 446–51 (1997).

70. B. Douglas Bernheim & Antonio Rangel, *Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers*, *in* ECONOMIC INSTITUTIONS AND BEHAVIORAL ECONOMICS (Peter Diamond & Hannu Vartiainen eds., forthcoming) (manuscript at 11, 26, *available at* http://www.stanford.edu/~bernheim/Behavioral%20 Public%20Economics%20Final.pdf).

is deceptively simple to designate one of them as "correct." When there are infinitely many discount factors, the selection is not nearly so simple. We suspect that advocates of the "mistake" interpretation have been misled by the quasi-hyperbolic approximation, which was originally adopted more for its mathematical tractability (relative to models of true hyperbolic discounting) than for its accuracy in describing human behavior.[71]

The presence of more than two discount factors raises the possibility that, unless the discount factor representing the highest degree of patience is always regarded as the appropriate standard, decision-makers can be too future-oriented as well as too impatient. They may fail to recognize that life is not forever and may not pluck enough flowers. Specifically, Kivetz and Keinan have shown in a number of studies that as temporal perspective lengthens, individual regret over the failure to seize the pleasures of life grows while guilty regret over indulgence falls, with the former ultimately predominating.[72]

A second attempt to justify using a longer-term rate as the normative rate is based on the idea that a planning rate is considered more than the acting rate. In other words, the planning rate is the result of a calm, collected, and thoughtful process while the acting rate is dominated by transient passions.[73] But this is far from the only plausible explanation of the difference between short and long-term discount factors. First, it is not unreasonable to believe that the *opportunity costs* of "virtue" become more evident as the moment of action arrives.[74] Far from being a less considered behavior, the present-oriented action may therefore actually be *more* considered and better informed. Second, Paul Glimcher and coauthors have found that a lower (more impatient) discount factor is applied *whenever* one of the outcomes compared is the earliest

71. *See* AINSLIE, *supra* note 25, at 210 n.29, 214 n.21; George-Marios Angeletos et al., *The Hyperbolic Consumption Model: Calibration, Simulation and Empirical Evaluation*, 15 J. OF ECON. PERSP. 47, 50 (2001).

72. This does not imply that the choice of a long-run benefit ("virtue") over a short-term indulgence ("vice") is always the source of predominant regret in the long run, but that it can be, especially when the optimal decision is not obvious. *See* Kivetz & Keinan, *supra* note 21, at 274.

73. Daniel Read, *Which Side Are You On? The Ethics of Self-Command*, 27 J. ECON. PSYCHOL. 681, 685 (2006). Read calls the planner the "pre-agent" and the actor "the agent." *Id.*

74. "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. 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." *Id.*

possible.[75] Specifically, agents apply the same discount factor for a choice between today and one day later as they do for a choice between sixty days from now and sixty-one days from now, *when the latter is the earliest possible option*.[76] This suggests that considered choice may not be at issue, because the same discount factor is applied even when the earliest possible date is two months away.

Third, the phenomenon of hyperbolic discounting may not be a strict matter of time-preference at all. Ariel Rubinstein has argued that differences in time periods seem more similar when agents contemplate them in the farther future than in the near term.[77] Goods delivered in 101 days or 111 days are more similar to each other than the same two goods delivered in 1 day or 11 days.[78] Under these circumstances the delay-attribute more or less drops out of consideration in the first case but not in the second. Thus the more patient "long-run" discount factor is the result of a relative failure to envision or appreciate future time delays.[79] This suggests that the short-term rate or rates might be more considered.

2. Framing and endowment effects

The framing problem is also evidenced by individuals making different choices for identical choice problems presented in different ways. Again, the inconsistency in choice allegedly reveals an underlying inconsistency of preferences. If we set aside that the different frames might actually matter to the individual's subjective wellbeing and suppose the frame really is irrelevant[80] we still have to ask: what are the true preferences? If the choice under Frame A corresponds to preference set A, and the choice under Frame B corresponds to preference set B, either A or B could represent the agent's true preferences. And here, too, the paternalist does not necessarily face a simple binary choice, as A and B might not represent the only way to frame the problem.

75. *See* Paul William Glimcher, Joseph Kable & Kenway Louie, *Neuroeconomic Studies of Impulsivity: Now or Just as Soon as Possible?*, AM. ECON. REV., May 2007, at 142.

76. *Id.* at 143–45.

77. Ariel Rubinstein, *"Economics and Psychology"? The Case of Hyperbolic Discounting*, 44 INT'L ECON. REV. 1207 (2003).

78. *Id.*

79. *Id.* at 1210.

80. Madrian and Shea argue that framing the 401(k) participation decision in such a way that enrollment is the default is likely to be seen by employees as "implicit advice" from employers who presumably know better. Brigitte C. Madrian & Dennis F. Shea, *The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior*, 116 Q. J. ECON. 1149, 1182 (2001). To the extent that this is the case, framing really does matter because it conveys information. *Id.*

Take the example of vacation time in employment contracts. Suppose that potential employees feel different, and thus negotiate differently, when more vacation time is a default part of the contract than when it is not. Under default A (less vacation time), added vacation time seems less valuable, so the employee does not strongly negotiate for it. Under default B (more vacation time), that added time seems more valuable, so the employee strongly resists its reduction.[81] This is a classic case of willingness-to-pay differing from willingness-to-accept, and thus evidence of internally inconsistent preferences. But which default rule corresponds to the agent's true preferences, representing his actual trade-off between leisure time and money? This question is crucial to the policy choice of an optimal default, since the wage rate will fall to compensate for longer vacation periods, yet the theory provides the policymaker with no means of choosing.[82]

3. Hot and cold states

The existence of a bias based on emotional states is supposedly revealed by an individual making different choices depending on whether he is in a "hot" or "cold" state. For instance, a person may choose to have sex or eat unhealthily when in a hot (aroused or hungry) state, yet refuse the same opportunity when in a cold (not-aroused or not-hungry) state. This seems to reveal an inconsistency—although this is likely a case when the emotional state itself has a large effect on the actual satisfaction gained from the activity. But let us suppose a real inconsistency is revealed here: the sex or junk food would be as physically satisfying either way. Emotional state A yields one choice, while emotional state B yields another. Still, which choice reflects the agent's true preferences?[83]

It might be that we could ascertain the true preferences on the basis of the absence or presence of subsequent regret. Consider that indulging in a pleasure as a result of a hot state may lead to a feeling of regret in

81. *See, e.g.*, Sunstein & Thaler, *Libertarian Paternalism Is Not an Oxymoron*, *supra* note 3, at 1176.

82. In cases like this, the tendency of the policymaker is to adopt an objective standard of welfare and set the default to the option that is "objectively" better. This constitutes an abandonment of the new paternalist project.

83. If we interpret this situation as a conflict of multiple selves (hot self/cold self) then taking sides is arbitrary. *See* B. Douglas Bernheim & Antonio Rangel, *Addiction and Cue-Triggered Decision Processes*, 94 AM. ECON. REV. 1558, 1572 (2004) ("Under that interpretation, our use of cold preferences as a welfare standard is arbitrary.").

the form of guilt—*but guilt itself is a hot state*.[84] It can and does pass. In the longer run, however, the individual may be relieved that he has not missed out on the pleasures of life. Which is the correct standpoint for the paternalist to adopt: the avoidance of immediate feelings of guilty regret after the indulgence, or the later avoidance of wistful regret over missed pleasures? The first presumes that the *initial* hot state (sexual arousal or hunger) distorts true preferences; the second presumes that the *subsequent* hot state (guilt) distorts true preferences. A reasonable case could be made for adopting either of these perspectives.[85]

Thus there are two preference sets to choose from, and again, no basis by which to choose except perhaps the paternalist's own preferences. And if we allow the existence of an interaction between the state of the agent during choice and his experience of the consequences of the choice, there may be more than two preference sets for the paternalist policymaker to choose from. As George Loewenstein recognizes,

> [I]t would clearly be suboptimal to make decisions that ignore visceral factors. Visceral factors do affect the marginal utility of different activities: eating is more pleasurable when one is hungry, and sex is more pleasurable when one is aroused. . . . Clearly, welfare maximization lies somewhere between the two extremes of making decisions that ignore visceral factors and treating visceral influences as no different from any other influence on tastes.[86]

Thus even if we were to assume unambiguous post-decision regret, this must be balanced against heightened enjoyment during or immediately following the decision. The paternalist policymaker is therefore faced with deciding the correct *balance* of the preferences corresponding to hot and cold states.[87]

84. *See* Kivetz & Keinan, *supra* note 21, at 280.

85. In the end, the application of the regret criterion is an empirical matter. Unfortunately, there has not been very much research on the pattern of regret consequences of actions.

86. Loewenstein, *Emotions in Economic Theory*, *supra* note 36, at 429.

87. The complexity of this problem has been recognized by Loewenstein and O'Donoghue, who recognize that no clear normative standard can come from this "dual-system" analysis. *See* George Loewenstein & Ted O'Donoghue, Animal Spirits: Affective and Deliberative Processes in Economic Behavior 19–21 (May 2005) (unpublished manuscript, *available at* http://www.arts.cornell.edu/econ/edo1/will.pdf). The paternalist must distinguish, on the one hand, the rational adaptation to the unconscious input of the affective system, that is, to the tacit personal knowledge of the kinds and sources of the individual's well-being, and, on the other hand, the "excessive" yielding to affective demands because of limited willpower. *Id.* at 38.

One possible response is to ask which preference set actually leads to greater long-run happiness.[88] Perhaps "cold" states typically lead to choices that produce more actual happiness. Here, the philosophical issues that we hoped to set aside in this article become impossible to bracket. How shall we measure actual happiness? Is it physical pleasure, as a hedonist would suggest? Or do the agent's other values also come into play? And if the latter—as seems most plausible to us—then how heavily should those values be weighed against physical pleasure (which is surely relevant even if not decisive)? Emotional states A and B may simply correspond to different relative weights attached to physical pleasure and other values—and again, theory gives us no means of determining which of these sets of weights, or which combination of these weights, corresponds to the agent's "true" preferences.[89]

In each case, the paternalist has to decide which among *equally viable* candidates to designate as the true preferences that will be privileged by policy.

VII. DISCOVERING THE EXTENT OF DECISION-MAKING BIAS

It is not enough to know that a bias exists. Nor is it enough to have identified a baseline of "true" preferences. The paternalist policymaker also needs to know the *extent* of the bias in order to design the appropriate solution to counteract it. This is not an easy task. Numerous problems arise in this process. For the remainder of this section, we will assume for the sake of argument that the problem (discussed in Part I) of identifying "true" preferences has somehow been resolved.

A. Lack of Precision in Measuring Extent of Bias

Precision in measuring the extent of any given bias has substantial policy relevance. A large bias will justify some policies that a small bias will not. The size of the bias will also affect the optimal degree of intervention. Excessive intervention can reduce welfare below pre-intervention levels. We will focus on four illustrative biases whose extent

88. Is the total undiscounted amount of lifetime happiness the relevant standard? On the other hand, if we must discount, then the issues of the previous section on discount rates reassert themselves.

89. One of the factors that enables Bernheim and Rangel to rationalize their use of cold preferences as the welfare standard is the *assumption* that the individual simply seeks to maximize discounted hedonic utility. *See* Bernheim & Rangel, *supra* note 83, at 1572 ("Since the individual has only one set of [true] preferences, discounted experiential utility . . . accurately measures his well-being, and is unambiguously the appropriate welfare standard.").

matters for policy: hyperbolic discounting, status quo bias, hot and cold states, and optimism bias.

1. Hyperbolic discounting.

In order to craft wise policies to correct problems created by present-bias, it is necessary to determine the extent of present-bias. If people need encouragement to save more, the extent of their present-bias will affect how much encouragement they require. The optimal size of a fat tax depends on the extent of present-bias in eating choices. Only after determining the extent of present-bias in the areas they wish to regulate could paternalists suggest a possible solution to counter present-bias.

Unfortunately, "[t]here is extraordinary variation across studies, and sometimes even within studies" in estimates of intertemporal discount factors.[90] Even when the same data set is analyzed using different, but standard, econometric techniques, there is often large variation in discount estimates.[91] Given the current technology of estimation, the "spectacular disagreement among dozens of studies"[92] implies an even greater variation in the predicted welfare effects of different policies based on "correcting" rates of excessive impatience. As we see below, even rather small differences in attributed discount factors can be associated with significant differences in welfare. We discuss two examples.

In an important Brookings study on savings for retirement, David Laibson and coauthors seek to evaluate the welfare impact of tax-deferred defined contribution retirement saving (DC) plans to consumers with self-control problems.[93] Such consumers tend to under-save because, although they recognize the "true" value of savings as between periods 2 and 3 and all other delayed pairs, they are excessively impatient as between the present and the next period (that is, they engage in quasi-hyperbolic discounting). As a consequence, individuals continually plan to save in the future, according to their true preferences,

90. Frederick, Loewenstein & O'Donoghue, *supra* note 22, at 393.

91. *See, e.g.*, *id.* at 385 (using an example with a range of discount rates between 1% and 14%).

92. *Id.* at 389. In addition, "there is no evidence of methodological progress in that the range of estimates does not seem to be shrinking with time." Dilip Soman et al., *The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones?*, 16 MARKETING LETTERS 347, 354 (2005).

93. David I. Laibson, Andrea Repetto & Jeremy Tobacman, *Self-Control and Saving for Retirement*, 1998 BROOKINGS PAPERS ON ECON. ACTIVITY 91.

but always fall short of their goals when the time arrives. Therefore, these individuals, if sophisticated enough to correctly forecast their lapses from optimal discounting, would value a commitment technology that would bind them to their plans, especially if it were costless and perfect. DC pension plans *approximate* such a technology because there are generally tax penalties for early withdrawal and because, if individuals change (lower) their contributions, the effect in increased consumption is somewhat delayed.[94] Laibson and coauthors provide several simulations that suggest significant differences in the welfare-enhancing effects of making DC plans available to individuals with different present-bias factors.[95] The present bias factor is typically represented by β, which is the *additional* discount applied to future periods when they are compared to the present. According to Laibson's calculations, the gross value of a DC plan to a twenty year-old high school graduate varies from 28% of his current annual consumption if $\beta = 1$ (that is, no present-bias), to 71% if $\beta = 0.85$, to 99% if $\beta = 0.8$.[96] And if $\beta = 0.60$, the impatience factor derived from much of the experimental evidence, the linear extrapolations of these simulations, suggest that "the true hyperbolic [excessive impatience] effect is two to three times as large as the effects reported above . . . ."[97] These results indicate that the desirability of policies to encourage more savings depends crucially on the extent of bias. Moreover, the appropriate amount of encouragement (i.e., how large the tax penalty for early withdrawal should be) will also depend on the extent of bias. Too much encouragement can cause a departure from the standard of true preferences that is as great as or greater than too little encouragement.[98]

94. *Id.* at 144–45.

95. *Id.* at 145–67.

96. *Id*. at 165.

97. *Id*.

98. Accurately ascertaining the extent of the impatience bias is also important in assessing the welfare impact of Social Security on naïve agents—that is, agents who have self-control problems of which they are unaware or which they forecast incorrectly. One argument for compulsory Social Security is that such individuals will, if left to themselves, save less than they "really" want and thus have a lower than optimal retirement income. Ayse İmrohoroğlu and coauthors conclude, based on their simulations, that Social Security does not raise welfare from the perspective of almost any age for individuals with impatience factors in the neighborhood of 0.85 to 0.90. Ayse İmrohoroğlu, Selahattin İmrohoroğlu & Douglas H. Joines, *Time-Inconsistent Preferences and Social Security*, 118 Q. J. ECON. 745, 781 (2003). This is because an unfunded retirement scheme, such as Social Security, lowers the aggregate capital stock and thus income at all ages. *Id*. at 770. While Social Security redistributes existing income to those in retirement, "the utility gains from increased old-age consumption are too small to offset the losses from reduced consumption earlier in life." *Id.* at 776. However, all this changes, as may be expected, when the

Now we turn our attention to optimal sin taxes. O'Donoghue and Rabin develop a model in which consumers choose between a composite good and a "sin good" defined as an immediately enjoyable good with longer-run bad health consequences.[99] O'Donoghue and Rabin simulate optimal taxes for plausible values of the other relevant parameters, as both the proportion of the population with self-control problems and the extent of their self-control problems vary. For example,

> [i]f half the population is fully self-controlled while the other half the population has a very small present bias of β=0.99, then the optimal tax is 5.15%. If instead the half the population with self-control problems has a somewhat larger present bias of β =0.90—which is still a smaller present bias (larger β) than often discussed in the literature—the optimal tax is 63.71%.[100]

Therefore, if the government were to estimate the present bias as the latter (lower β) when in fact it was the former, it would *reduce* consumer welfare by imposing a tax about twelve times too large. Notice also that O'Donoghue and Rabin's approach includes assumptions about what percentage of the population is afflicted by present-bias[101]—an issue we will address more fully later.

The difficulty for policy prescriptions is that no one is very confident about the true impatience or present-bias factor (β), nor about the proportion of the population subject to the bias. This is one reason that, in the various studies discussed above, the authors show the effects on welfare for various calibrations of the relevant parameters. In sum, the extent of the impatience bias is very significant in determining whether a specific paternalist policy increases or decreases welfare relative to the status-quo. *Current estimates are unable to provide a basis for policy prescriptions that reliably increase welfare*. At best, policies derived from the current state of knowledge can only produce certain *objective results*, like more saving or lower junk-food consumption, that may or may not increase welfare. Therefore, the new paternalism, supposedly

degree of impatience increases. Under those circumstances, the amount of under-saving may be so great that the increase in income during retirement brought about by Social Security payments will swamp the effects of a lower aggregate capital stock. In fact, the simulations reveal that "[s]ocial security does significantly raise welfare with β = 0.60 . . . ." *Id*. at 781. Obviously, government policies regarding savings for retirement will be affected substantially by the extent of the impatience bias. A relatively small bias may suggest the substitution, in whole or part, of fully-funded or private retirement plans for the current Social Security scheme.

99. O'Donoghue & Rabin, *Optimal Sin Taxes*, *supra* note 3.

100. *Id.* at 1838.

101. *Id.*

based on the underlying normative preferences of individuals, shades into the old paternalism, based on what is "objectively best" in the opinion of an outside observer.

2. Status quo bias

If default savings plans are justified on grounds of status quo bias, then we need a measure of the extent of that bias. The greater the status quo bias, the more the selected default savings plan matters because more people will stay with it longer.[102] If that plan is not optimal, then individuals will be stuck in a relatively low-welfare savings outcome. How bad this situation is and how long people will be in it depend on the welfare losses from suboptimal savings and the extent of the bias.[103]

In their seminal study, Samuelson and Zeckhauser found that the status-quo bias, even where statistically significant, differs in size across tasks and alternatives—from substantial effects to small effects.[104] Whether this is due to systematic contextual factors, the inherent variability of the phenomenon, or difficulties in measurement technique, is impossible to say at this time. Furthermore, whether these effects, regardless of their magnitude, are caused by rational transaction cost factors or behavioral biases is difficult to determine.[105]

Moreover, Samuelson and Zeckhauser found that the bias is larger when the individual's preference for a neutrally-presented alternative to the status-quo is weaker.[106] Thus, the size of the bias is likely to depend on the default; if people know they are already saving something by

102. Status-quo bias is usually estimated by the (disproportionate) frequency with which the status-quo option is accepted by decision-makers. The duration of the bias—how long people stay with the option—has not been systematically measured.

103. The issue here is somewhat more complex because what we must know is optimal 401(k) savings, since people save in other ways. This is only indirectly related to the general rate of excessive impatience.

104. Samuelson & Zeckhauser, *supra* note 30, at 15–17 tbls.1a, 1b & 1c. The absolute size of the status-quo bias is SQ-NEUT and the relative size is (SQ-NEUT)/NEUT where SQ is the choice frequency for a given alternative when it is in the status-quo position and NEUT is the frequency when the alternative is presented neutrally. *Id.* at 15–17.

105. In their analysis of the impact of status-quo bias on decisions regarding enrollment in 401(k) programs, Madrian and Shea observe, "Unfortunately, there is no way to disentangle the magnitude of rational, transaction costs motivated procrastination from behavioral, self-control motivated procrastination in the data." Madrian & Shea, *supra* note 80, at 1180. They do note, however, that there is a "possibility of the latter." *Id.*

106. They state their equivalent conclusion in terms of the converse proposition: "The stronger was an individual's preference for a selected alternative, the weaker was the bias." Samuelson & Zeckhauser, *supra* note 30, at 8. *See also* James J. Choi, David Laibson, Brigitte Madrian & Andrew Metrick, *Optimal Defaults*, AM. ECON. REV., May 2003, at 180, 183–84.

default, they may be less likely to take the time to change to a better plan. Because of the generally low returns to the default allocations, Choi and coauthors found that automatic enrollment produced offsetting effects: "While higher participation rates promote wealth accumulation, the low default savings rate and the conservative default investment fund undercut accumulation," and, in their sample, the two effects were approximately equal in magnitude.[107] So, in the aggregate, these individuals were in no better position than before. On the other hand, the farther the default is from the individual's optimum savings rate, the greater the probability that he will opt out of the default and begin saving optimally. A sufficiently inappropriate default will weaken the status-quo bias and motivate change.[108] In this case, the paternalist should not be searching for a welfare-enhancing default, but for one that is far enough from it to encourage active decision-making. Therefore, the nature of the paternalist's task will depend not only on knowing the relevant size, persistence, and cause of the status-quo bias, but also its responsiveness to the alternatives considered. Neither the economist nor the paternalist has adequate measures of any of these factors.[109]

3. Endowment effects

The size of the endowment effect clearly determines whether any paternalistic change in the assignment of default contractual rights can increase welfare. If endowment effects are weak or even nonexistent, then *even if the paternalist selects the optimum rights package*, no

107. James J. Choi, David Laibson, Brigitte Madrian & Andrew Metrick, *For Better or For Worse: Default Effects and 401(k) Savings Behavior* 2 (Pension Research Council, Working Paper No. 2002-02, 2002).

108. *See* Choi, Laibson, Madrian & Metrick, *Optimal Defaults*, *supra* note 106, at 183–84; *see also* Samuelson & Zeckhauser, *supra* note 30, at 8 ("The stronger was an individual's preference for a selected alternative, the weaker was the bias.").

109. Another superficially attractive possibility is to choose a default that minimizes the total realized costs of opting out. However, this is not a welfare maximizing or enhancing standard in the presence of status-quo bias. As we have seen above, a default that motivates people to abandon more rapidly their suboptimal savings rate may be a good thing. This implies that the correct standard is the minimization of the *sum* of the realized costs of opting out *and* the flow losses due to too little or too much savings. In other words, higher realized costs of opting out would in fact be welfare enhancing if they were accompanied by a larger reduction in the costs of nonoptimal savings. *But see* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 109 (2008) (claiming that a low rate of opting out under an automatic enrollment default is welfare-enhancing because it reveals that people are in a better position). Individuals could stay in that position simply because they have not been sufficiently motivated to choose a more nearly optimal savings plan, that is, because they are experiencing the very status-quo bias that Thaler and Sunstein view as an important cause of suboptimal savings.

purpose is served by presuming vacation time, dismissal only for cause, etc., in employee contracts beyond saving on transaction costs. Of course, if the paternalist does *not* select the optimum rights package, transactions costs will be increased.

Until recently, the behavioral literature accepted the existence of endowment effects without much controversy. Surprisingly, the existence of these effects has never been adequately tested. Kathryn Zeiler and Charles Plott have undertaken and reported experiments that control for the most important factors that may be responsible for the appearance of endowment effects.[110] When factors that could plausibly affect the nature of the good in question (such as whether it is perceived as a gift from the experimenter which could be impolite to exchange, or whether the endowment is really a signal of private information from the experimenter or perhaps other subjects) are eliminated, the results suggest "[e]ither no 'endowment effect' of the sort predicted by prospect theory exists [in these experiments] or the effect is sufficiently weak that other phenomena easily swamp it."[111] But notwithstanding these results, let us suppose that true endowment effects exist. Accurate measurement of their magnitude will determine the efficacy of new default rules in improving welfare. If the effects are small, as Zeiler and Plott suggest, then default rules will simply increase transaction costs for people to return to their favored packages. The more difficult the default is to escape, the greater will be the resulting loss.

4. Hot and cold states

To create an optimal policy justified on the basis of hot-state bias, such as a cooling-off period, policymakers need to know how much people are affected by their hot states; that is, to what extent their decisions are distorted. It is possible they may not be distorted at all. The process by which they are supposedly distorted is through the "empathy gap"—the tendency of individuals in a hot or excited emotional state to overestimate the intensity of the hedonic consequences of an event or

110. Charles R. Plott & Kathryn Zeiler, *Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?*, 97 AM. ECON. REV. 1449 (2007). This study stands out as the most rigorous attempt to date to control for confounding factors. *See id*. at 1454–56.

111. *Id.* at 1463.

good later on when the hot state has dissipated.[112] However, hedonic consequences are not the sole, and, in many cases, not even the primary determinants of choice, as when people sacrifice personal pleasure to send their children to college or to pursue some form of excellence. Thus, overestimation of hedonic consequences may not have a significant impact on many decisions.

Furthermore, hotness and coldness refer to the state of the affective system. But even this is a simplification of the problem, because affect is not just hot or cold. Affect is an important element in the decision-making process; it comprises the individual's motivational system—hunger, thirst, and the desire to have sex are just a few examples. The affective system suggests the options an individual must consider. Then it reduces or flags these options in accordance with the individual's specific goals, time frames, and perhaps most importantly, his acquired knowledge of local circumstances.[113] For example, a pleasant or unpleasant feeling may follow from thinking about meeting an old acquaintance or thinking about consuming a product with particular mental associations. These images are marked by a somatic state.[114] So an option that is likely to result in such a meeting or consuming such a product may immediately, that is, without much deliberation, be chosen or discarded by the individual as a result of the somatic state. Consequently, a rational choice can *seem*, at times, as if it were choice without reasoning. We might see an individual jumping to conclusions, hastily eliminating alternatives and making decisions, and "wanting" things without a "sufficient" reason.[115] None of this suggests that choices are distorted; this is the way human beings choose—a combination of explicit deliberation and affect. Affect is part of rationality.

112. *See* Timothy D. Wilson & Daniel T. Gilbert, *Affective Forecasting*, 35 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 345, 365–66 (2003). It is also the case that when in a cold state people underestimate the intensity of feelings in a hot state. *Id*.

113. ANTONIO R. DAMASIO, DESCARTES' ERROR: EMOTION, REASON AND THE HUMAN BRAIN 173, 181–83 (1994).

114. The body states produced by processes of the affective system necessary for rational decision-making may be conscious or unconscious, that is, they may or may not constitute *feelings*. Body states may be activated by stimuli but not be the focus of awareness or attention. Nevertheless, they can affect "cognitive processes in a covert manner and thus influence the reasoning and decision-making mode." *Id.* at 185.

115. This is particularly likely in the case in which the individual himself will not experience a feeling of liking, hating, fearing, and so forth. He will simply approach or avoid, want or not want, without explicit liking or disliking.

Of course, it is possible for the affective system to break down and to produce distorted choices.[116] If we allow that this will happen in some cases, what must the policymaker know? First, he must be able to distinguish distortive affect from normally-functioning affect. As we have seen, appearances do not suffice. Second, even if the policymaker knows that a certain type of affective state can be distortive, he must know to what extent it distorts in a particular context. Small distortions are hardly worth policy attention. Third, he must know something about the rate at which this particular distortive hot state dissipates. Cooling-off periods should be calibrated to this rate. A longer cooling-off period would impose excess costs on the sellers and then to the consumers. And, finally, he must know the degree to which ex post rationalization of a decision that cannot be changed will obviate the need for cooling-off periods.[117]

5. Optimism and availability bias

To create an efficient policy designed to counteract optimism bias, such as requiring risk narratives for risky products or investments, policymakers need to know both the extent of people's excessive optimism (how much they underestimate the risk) and the extent of their availability bias (how much they respond to a narrative with varying levels of scariness). More excessive optimism points toward more and scarier narratives, greater availability bias toward fewer and less scary narratives. But what is the standard by which the paternalist's policy will be judged? Presumably, he wants the two biases to balance at the point where people make the rational decision. It would not be sufficient to know the "correct" estimate of risk (and then to try to induce this perception by the appropriate narrative), because how much risk a person ought to bear is independent of neither his subjective attitude toward risk

116. *See, e.g.*, Kent C. Berridge, *Pleasure, Unfelt Affect, and Irrational Desire*, *in* FEELINGS AND EMOTIONS: THE AMSTERDAM SYMPOSIUM 243, 254–59 (A.S.R. Manstead, N.H. Frijda & A.H. Fischer eds., 2004) (using the example of irrational choice arising from addictions).

117. Wilson and Gilbert argue that cooling-off periods might actually make people less satisfied with their decisions because they inhibit the process of rationalization. Timothy D. Wilson & Daniel T. Gilbert, *Affective Forecasting: Knowing What to Want*, 14 CURRENT DIRECTIONS IN PSYCHOL. SCI. 131, 133 (2005) ("When people make a decision that is difficult to reverse, such as buying a sweater from a store with a 'no returns' policy, they are strongly motivated to rationalize the decision and make the best of it. When people can more easily undo a decision, such as buying a sweater they can return, they are less motivated to rationalize their choice, because they can always change their minds. *Consequently people are often happier with irrevocable choices* because they do the psychological work necessary to rationalize what they can't undo." (emphasis added)).

nor his subjective assessment of the cost of the bad outcome associated with it. The paternalist needs some indication of, say, the right product to buy or the right mutual fund in which to invest. However, if he knew this, then he would not need to worry about offsetting optimism bias with appropriate risk narratives, because policy could simply command the correct outcome.

B. Absence of a Single Measure of Bias, Even Intrapersonally

To make matters more difficult, a *single* measure of any given bias generally does not exist, even for a single individual. Different degrees of bias will exist depending on the choice situation. Here we focus on hyperbolic discounting and hot-and-cold states.

1. Hyperbolic discounting

As discussed earlier, people's actual behavior in situations of intertemporal choice appears to approximate hyperbolic (not quasi-hyperbolic) discounting. This means that there is no single factor β that represents the agent's degree of present-bias. Instead, the extent of bias depends on the distance between the two future periods compared and the present.

Suppose the paternalist policymaker has (somehow) determined that the agent's true preferences are best represented by some fixed discount factor δ. If the agent's actual behavior approximates hyperbolic discounting, then the agent will discount the future too much when comparing periods relatively close to the present. But what is the extent of that bias? The answer will depend on how close the two periods compared are to the present. The closer they are to the present, the greater will be the present-bias. On the other hand, the agent will also discount the future too little when comparing periods sufficiently far from the present. This is a necessary result of the paternalist's having designated a fixed discount factor δ as correct, when actual behavior reflects discount factors both higher and lower than δ. This conclusion could only be avoided by the paternalist having assumed the correct discount factor is the highest (most patient) one the agent ever exhibits.

The implication for policy is that bias-correcting policies should be calibrated to the distance from the present of the intertemporal decisions being made. Take, for instance, a fat tax designed to curb junk-food consumption. The tax would need to be higher when a person is buying food for immediate consumption—say, at a restaurant or convenience store. The tax would need to be lower when a person is buying food for

more distant consumption—say, at a grocery store. And depending on the policymaker's judgment about the correct amount of discounting, it might even be necessary to subsidize fat, rather than taxing it, for very distant consumption—say, when planning for a celebration a year from now (like a wedding). Again, this follows directly from the selection of a single correct discount factor in the context of hyperbolic discounting. Even if the policymaker has selected an extremely high discount factor so that no subsidies are required, he still needs to make the tax a function of the degree of present-bias that applies to any given time frame, which requires the policymaker to have knowledge of that present-bias.

Obviously this is impractical. In reality, the policymaker would most likely adopt a single tax that would apply regardless of context or time frame. Such a tax would yield problems of both under-correction (the tax would be too low for decisions close to the present) and over-correction (the tax would be too high for decisions far from the present). Since any change in the tax will tend to produce more of one problem and less of the other, the policymaker will have to weigh these effects against each other to decide the best tax—and again, that requires having knowledge of the actual extent of present-bias for different time frames.

2. Hot and cold states

The impact of the hot-state bias on decision-making depends on the intensity of the relevant visceral factors.[118] There are degrees of anger, fear, hunger, or sexual desire. It stands to reason that more intense emotions will distort decisions more than do less intense ones.[119] The degree or intensity of visceral factors depends on the context of the decision. No decision defined in objective terms, such as whether to marry or buy a car, is necessarily a hot decision. Whether it is and whether it produces suboptimal choices "depends on a wide range of influences."[120] These include "how recently a drive was satisfied and on the presence of arousing stimuli" as well as "the interaction of situational factors and construal processes and on internal psychobiological factors."[121] Although some types of decisions are no doubt more likely to be affected by visceral states, there will be significant intra-individual

118. George Loewenstein, *Out of Control: Visceral Influences on Behavior*, 65 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 272, 273 (1996).

119. *Id.*

120. *Id.* at 281.

121. *Id.*

variation depending on the particular context. As a result, waiting periods for decisions to take effect or options to allow people to revoke prior decisions will have different consequences as the context varies. The net consequences may be costly in some circumstances and beneficial in others.[122] Whether a general rule, even adapted to particular *types* of situations, is beneficial *overall* requires knowledge of the relative frequency of the relevant contextual factors—knowledge no one has. Furthermore, the presumably different rates of hot-state dissipation (derived from the differential presence of contextual factors) will also determine the optimal length of the cooling-off period.[123]

VIII. ACCOUNTING FOR SELF-DEBIASING

People have numerous means at their disposal to mitigate the effects of their own biases. We will refer to these methods as "self-debiasing" or "self-regulation." We do not claim, however, that self-regulation effectively eliminates all or even most biases. Our argument is rather that the existence of such methods implies that some paternalistic policies that appear desirable at first blush are either unnecessary or in need of softening (lower sin taxes, shorter cooling-off periods, etc.) to account for the *extent* to which the biases have already been addressed privately. Policy measures that do not take account of self-debiasing can move the individual even farther away from his optimal decision than he would be in the absence of such policies.

A. The Many Varieties of Self-Debiasing and Self-Regulation

The most obvious form of self-regulation is simply the exertion of willpower. But in an important sense, willpower comes into play too late. When the individual is already exposed to a temptation, direct resistance can be very costly. Individuals, however, are more inventive about the methods they choose to achieve their long-run objectives. Self-regulation "consists of a wide range of cognitive and motivational operations, such as acting quickly to take opportunities, ignoring distractions, acting flexibly in response to situations, overcoming obstacles, and managing

122. In those cases in which the distortive aspect of hot decisions is small, the delay or option-to-revoke costs will outweigh the benefits.

123. Sometimes the hot state is caused by contemplation of the decision itself such as those relative to death, disease, accidents, and terrorism. Therefore, the hot state will not dissipate so long as the decision is ultimately made. In these cases neither delay nor option-to-revoke seems to have any paternalistic benefits. So the net result of having such cooling-off periods is costly. *See* Jeffrey A. Blumenthal, *Emotional Paternalism*, 35 FLA. ST. U. L. REV. 1, 61–62 (2007).

conflicts between goals."[124] More specifically, self-regulation functions to reduce the impact of behavioral biases by using strategies that are cognitive, environmental, and directly behavioral.[125] Cognitive strategies include focusing on the benefits of reaching one's goal, distracting oneself from undesirable behavior by using imagery of better alternatives, and using self-praise to commend oneself for achieving an important goal.[126] Environmental strategies include avoiding people, situations, settings, and even times of day when temptations are strong.[127] Directly behavioral strategies include increasing social support, utilizing cues about one's important goals, rewarding oneself for desirable behavior or punishing oneself for undesirable behavior, and creating ways to make desirable behavior itself more enjoyable.[128]

The following is a partial list of debiasing strategies in each general category. The large variety of these strategies and their connection to particular circumstances of time and place should make obvious the scope of the difficulties paternalists face in trying to account for them in the determination of welfare-enhancing policies.

1. Cognitive strategies

a. Resolutions and commitments. These mental devices focus a person's attention on those situations and choices in which his own biases are most likely to be manifested. A person who suffers from weakness of will when it comes to eating might make a resolution never to eat desserts except on special occasions. A person with a marked tendency to make rash decisions in hot emotional states—say, when confronted with the opportunity to commit adultery—might resolve to physically remove himself from the situation before making any decision or to count slowly to ten before taking any action.

b. Mental accounts and budgets. As a means of setting boundaries and reminding themselves of their resolutions, people sometimes adopt

124. Gráinne M. Fitzsimons & John A. Bargh, *Automatic Self-Regulation*, *in* HANDBOOK OF SELF-REGULATION: RESEARCH, THEORY AND APPLICATIONS 151, 151–52 (Roy F. Baumeister & Kathleen D. Vohs eds., 2004) (citing P.M. Gollwitzer & G.B. Moskowitz, *Goal Effects on Action and Cognition*, *in* SOCIAL PSYCHOLOGY: HANDBOOK OF BASIC PRINCIPLES 361, 368 (E.T. Higgins & A.W. Kruglanski eds., 1996)).

125. For a list of self-regulatory strategies, see ENCYCLOPEDIA OF MENTAL DISORDERS, *Self-Control Strategies*, http://www.minddisorders.com/Py-Z/Self-control-strategies.html.

126. *Id.*

127. *Id.*

128. *Id.*

mental accounting devices to keep certain behaviors within limits.[129] For instance, a person might establish an entertainment budget; he allows himself to spend as much as he likes on entertainment up to a chosen limit, but will not let himself exceed it. Or he might establish a fund for household expenses that cannot be tapped for other purposes. Mental budgets can enable indulgence as well as limit it, such as when someone commits himself to take a vacation (perhaps to overcome a tendency toward overworking).

2. Environmental strategies

a. Submission to social controls. These are efforts to enlist outsiders to assist in the keeping of one's commitments. Someone trying to quit smoking may advertise that intention to friends and family, so they will remind him of his commitment and frown on deviations from it. Formal organizations like Alcoholics Anonymous and Weight Watchers play the same role, providing a support network that lowers the cost of following commitments and raises the cost of breaking them. Strotz provides the more extreme examples of getting married "for the sake of 'settling down'" or joining the army as methods of precommitting financial or economic actions.[130]

b. Self-constraining devices. These devices structure the external environment to raise the cost of some activities and lower the cost of others. People trying to quit smoking sometimes throw away their cigarettes to remove the temptation. People with eating problems may refuse to allow especially tempting foods in their home. People who have difficulty saving can opt to have automatic monthly transfers from their checking accounts to their savings accounts. Gamblers can limit their opportunities to exceed self-imposed limits by leaving their credit and ATM cards at home.

3. Directly behavioral strategies

a. Internal rewards and punishments. It is not uncommon for people to affect their choices by means of internally imposed incentive schemes,

129. That people use mental budgeting to control their behavior is well established. *See, e.g.*, Chip Heath & Jack B. Soll, *Mental Budgeting and Consumer Decisions*, 23 J. CONSUMER RES. 40 (1996); Richard Thaler, *Mental Accounting and Consumer Choice*, 4 MARKETING SCI. 199 (1985); Klaus Wertenbroch, *Consumption Self-Control by Rationing Purchase Quantities of Virtues and Vice*, 17 MARKETING SCI. 317 (1998).

130. Strotz, *supra* note 25, at 173.

by which they give themselves rewards for more favored behavior and punishments for less favored behavior. For example, someone trying to lose weight might reward herself for meeting weight-loss goals with permission to go to a movie or buy another music CD. The phenomenon of self-gifting has been documented in a series of papers,[131] and the efficacy of self-reward schemes in motivating greater effort and performance has also been shown.[132]

B. The Significance of Context for Self-Regulation

Consideration of the various methods used by real people to regulate their own behavior reveals the overriding importance of context.[133] Resolutions, commitments, mental budgets, and internal rewards and punishments typically depend for their application on specific features of time and place: what time of day it is; whether one is at work, at home, or on vacation; whether the present situation is a special occasion like a birthday or wedding; and so on. Self-constraining devices and submission to social controls are devices designed to *affect* one's context by inserting costs, benefits, barriers, and reminders that would not otherwise be present.

1. Self-regulation in the laboratory versus in the wild

As has been said, the existence of self-debiasing measures does not mean that all biases are perfectly corrected. Individuals may not know all of their biases, and the methods they adopt may not succeed—or may succeed too well, as in the case of anorexics or tightwads. The point is

131. *See* David Glen Mick, *Self-Gifts*, *in* GIFT-GIVING: A RESEARCH ANTHOLOGY 99 (Cele Otnes & Richard F. Beltramini eds., 1996); David Glen Mick & Michelle DeMoss, *Self-Gifts: Phenomenological Insights from Four Contexts*, 17 J. CONSUMER RES. 322 (1990).

132. *See* Albert Bandura & Dale H. Schunk, *Cultivating Competence, Self-Efficacy and Intrinsic Interest through Proximal Self-Motivation*, 41 J. PERSONALITY & SOC. PSYCHOL. 586, 586–87, 595–97 (1981); Albert Bandura & Bernard Perloff, *Relative Efficacy of Self-Monitored and Externally Imposed Reinforcement Systems*, 7 J. PERSONALITY & SOC. PSYCHOL. 111, 111, 114–116 (1967).

133. All actions derive their meaning from context. Consider the simple act of an individual touching his nose with a finger. The meaning changes with the context in which it takes place. If the individual is being asked to touch his nose as part of an experiment, the purpose of which is unknown to him, then he may see the act as simply obeying the instructions for the sake of science or for some payment. If it is a part of a neurological exam, then the context is the health of the individual or the diagnosis of a possible disease. Alternatively, it may simply be scratching an itch or swatting a fly. Or it may be a socially-recognized gesture of disapproval. Context determines meaning. For a detailed analysis, see Shaun Gallagher & Anthony J. Marcel, *The Self in Contextualized Action*, 6 J. CONSCIOUSNESS STUD. 4 (1999).

that any debiasing policy will only be successful to the extent that it takes self-debiasing efforts into account. Measuring the extent of bias absent self-debiasing efforts will overstate the degree of bias realized in behavior, and hence the amount of correction required. Without a clear and realististic context, laboratory measurement of self-regulation or control will not capture this.

To see why this should matter to the paternalist policymaker, let us take the case of an individual who faces future consequences to be balanced against current costs or benefits. His actual rate of impatience is greater, we assume, than what would be dictated by his "true" preferences.[134] Let us suppose, charitably, that the policymaker already knows the individual's true time preference, as well as the unmodified extent of his present-bias (that is, excessive impatience). Nevertheless, he still needs to know the degree to which the individual's self-regulatory mechanisms counteract his own impatience. This would give the policymaker an *effective* or *operational* level of present-bias, which will—if the person's self-regulation works at all—differ from his unmodified present-bias. This knowledge is necessary to determine, for example, the rate of taxation on present benefits that will lower the effective rate of impatience to the correct level.

To solve the problem, the paternalist needs to measure the amount of self-debiasing that occurs, but doing so is inherently problematic. Much of our evidence of decision-making biases derives from laboratory experiments. But laboratory experiments cannot capture all debiasing efforts, because self-debiasing efforts are context-dependent: the person with a weight problem resists desserts *except on special occasions*, the person trying to save more money signs up for automatic deductions from *paychecks* (but not unexpected windfalls), etc. Lab environments, on the other hand, are typically devoid of context. Even if the experiment designer deliberately structures the experiment to create the illusion of

134. Not all self-regulation or "self-control" problems are impatience problems. There can be a negative difference between the actual rate and the normative rate. *See, e.g.*, John Ameriks, Andrew Caplin, John Leahy & Tom Tyler, *Measuring Self-Control* 1 (Nat'l Bureau of Econ. Research, Working Paper No. 10514, 2004) ("The current view of self-control problems as involving the need to suppress the immediate urge to consume is inadequate. In our sample, 'present-bias' (the urge to consume today more than would be ideal) is no more prevalent than is 'future-bias' (a tendency to consume less today than would be ideal) . . . ."). The individual can have "excessive patience," as when he operates under the comforting illusion that he will not die or grow infirm or be less capable over time of enjoying physical activity. *See* Wojcieh Kopczuk & Joel Slemrod, *Denial of Death and Economic Behavior*, 5 ADVANCES IN THEORETICAL ECON., Jan. 2005, art. 5, at 2–4, *available at* http://www. bepress.com/bejte/advances/vol5/iss1/art5. We ignore this here for purely heuristic reasons, but caution that this hyperopia can complicate policymakers' decisions.

context, this effort cannot capture self-debiasing efforts that seek to achieve *overall* outcomes by differing *across* contexts. What may appear to be a bias in a particular context could be part of an overall plan that creates a deliberate exception in that area. The strategy of eating dessert only on special occasions, for example, rations fat consumption by defining narrow contexts in which it is allowed. A lab environment can duplicate one context, but not all the contexts relevant to the individual's overall strategy.

The context may be inferred by external observers of situations when observing human behavior "in the wild," or it may be supplied by the observers in experimental situations. The main question for us is whether the context supplied by observers or inferred by subjects in an experiment is equivalent to the context of the real-world situations to which the results of these experiments are generalized.[135] Thus, if an individual is asked as part of an experiment whether he prefers a larger, later reward as opposed to an earlier, smaller one over various intervals of time, he may or may not display a range of time-discounting propensities that reflect his real-world behavior. This will depend on the degree of similarity in context between the experiment and the wild.

Thus, self-regulation is context-dependent. The drive for generality in experiments, on the other hand, usually produces minimal or antiseptic context such as designating some actors as buyers or sellers, determining the set of alternatives, and ordering their presentation.[136] The experimenters may worry that providing "too much" context limits the applicability of results to the real world. But precisely the opposite is the case if the purpose is to inform paternalist policy. To capture the real world of choice, we must see choices in their self-regulatory context.

135. The equivalence of context has implications for both problem construal ("What am I being asked to do?") and for the nature of the solution ("How should I behave?"). *See, e.g.*, Glenn W. Harrison & E. Elisabet Rutström, *Doing It Both Ways—Experimental Practice and Heuristic Context*, 24 BEHAV. & BRAIN SCI. 413, 413–14 (2001) ("Field referents can often help subjects overcome confusion about the task. . . . [Even] [i]n cases where the subject understands all the relevant aspects of the abstract game, problems may arise due to the triggering of different methods for solving the decision problem. The use of field referents could trigger the use of specific heuristics from the field to solve the specific problem in the lab, which otherwise may have been solved less efficiently . . . ." (citations omitted)).

136. *See* George Loewenstein, *Experimental Economics from the Vantage-Point of Behavioural Economics*, ECON. J., Feb. 1999, at F25, F29 (1999) ("Many experimental economists seem to view their enterprise as akin to silicon chip production. Subjects are removed from all familiar contextual cues. . . . [B]uyers and sellers become 'persons A and B,' and all other information that might make the situation familiar and provide a clue about how to behave is removed.").

There is no such thing as general or abstract self-regulation[137]—although the contextual nature of self-regulation is obscured by the popular idea that it is derived from some homogeneous source such as inner strength or willpower.[138] Thinking in this way is apt simply to result in measuring choice propensities in under-defined contexts.

2. The automaticity of unconscious self-regulation

There is important and growing evidence suggesting that "conscious processes are neither necessary or even typical for effective self-regulation . . . ."[139] Much self-regulation must be non-conscious to be effective in view of the limited capacity of individuals to deal with a complex and rapidly-changing environment in a fully deliberative manner.[140] There are, for example, unconscious processes associated with selective attention, that is, the focusing on important or superordinate goals.[141] There is also unconscious modulation of emotional states that might threaten the attainment of these goals.[142] And subliminally activated goals have been shown to "guide behavior in a purposive, though nonconscious, manner . . . ."[143] These unconscious processes are triggered by the local and personal circumstances of the individual, that is, by his self-regulatory context, to an even greater extent than the conscious processes discussed above.[144] For example, simply thinking about people with whom one has a relationship, such as family, friends, and colleagues, can automatically "activate goals that

137. It is possible to make an even broader claim. *See id.* at F30 ("A major discovery of cognitive psychology is the degree to which all forms of thinking and problem solving are context-dependent . . . .").

138. This does not seem to be the case (or, at least, the metaphor does not seem appropriate) since there are *intra*personal differences across self-regulatory tasks and various situations. *See* Daniel Cervone, *People Who Fail at Self-Regulation: What Should We Think of Them—and How?*, 7 PSYCHOL. INQUIRY 40, 41 (1996).

139. Fitzsimons & Bargh, *supra* note 124, at 151 (providing a partial survey of the relevant literature).

140. *Id.* at 152.

141. *Id.*

142. The existence of mood regulation tends to counteract "irrational" pressures on decision-making when the stakes are high. *See* Ralph Erber, Maureen Wang Erber & Jennifer Poe, *Mood Regulation and Decision-Making: Is Irrational Exuberance Really a Problem?*, *in* 2 PSYCHOL. OF ECON. DECISIONS: REASONS & CHOICES 197, 204–05 (Isabelle Brocas & Juan D. Carrillo eds., 2003).

143. Fitzsimons & Bargh, *supra* note 124, at 153–55 (citation omitted).

144. *See id.* at 156–57 (discussing how social environment and personal relationships can affect unconscious self regulation).

guide and regulate the self's actions in a given situation . . . ."[145] These goals are generally those congruent with the attitudes of the others.[146]

Social norms relevant to the particular environment in which the individual acts are also sources of automatic processes.[147] Most interestingly, an automatic form of self-control known as "counteractive self-control" can be triggered by *imagining* the temptation. This is a proactive or ex ante adjustment of the relevant choice variables.[148] Counteractive self-control may involve changing the "objective" choice situation by self-imposing a penalty for the failure to achieve one's long-term goal.[149] Additionally, it may change the psychological meaning of the choice situation by raising the subjective value of the long-term goals and decreasing the subjective aversion of the short-term costs.[150]

Unconscious self-regulation is not easily observable. The target agent himself is unaware of its operation. It may take quite subtle forms, as we have seen above. For the new paternalist, accounting for this type of self-regulation is thus particularly difficult.

IX. ACCOUNTING FOR INTERDEPENDENT BIASES

The simultaneous existence of more than one bias affecting the individual's cognition or behavior poses a difficult problem for policy choices grounded in the new paternalism.[151] Almost universally, in the current state of research, only one bias at a time is studied.[152] But since we have good reason to believe that simultaneous biases are likely,

145. *Id.* at 157 (citations omitted).

146. *Id.*

147. *Id.* at 156.

148. Thus, counteractive self-control is not dissonance reduction. *See* Ayelet Fishbach & Yaacov Trope, *The Substitutability of External Control and Self-Control*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 256, 259 (2005) ("CCT [Counteractive Control Theory] concerns proactive attempts to enact what one ideally prefers, whereas dissonance concerns attempts to reduce the discomfort produced by having failed to enact what one prefers.").

149. *See id.*

150. For a survey of results, see Yaacov Trope & Ayelet Fishbach, *Going Beyond the Motivation Given: Self-Control and Situation Control over Behavior*, *in* THE NEW UNCONSCIOUS 537, 537–51 (Ran R. Hassin, James S. Uleman & John A. Bargh eds., 2005).

151. This problem should be distinguished from those arising from the existence of multiple biases within a population. We do not deal with this here.

152. *See* Hanming Fang & Dan Silverman, *Distinguishing Between Cognitive Biases*, *in* BEHAVIORAL PUBLIC FINANCE 47, 48 (Edward J. McCaffery & Joel Slemrod eds., 2006) ("So far, both the theoretical and the empirical studies in economics have tended to investigate the implications of cognitive biases and heuristics one bias at a time . . . .").

merely finding a bias that is significant both statistically and in size is not sufficient to conclude that the associated behavior is suboptimal.[153]

The identification of a myriad of cognitive and behavioral biases across hundreds of studies, as well as sometimes the identification of more than one bias within a single study, is good prima facie evidence that individuals are subject to multiple biases. Joachim Krueger and David Funder present a "partial list" of forty-two cognitive biases, including numerous opposite or contradictory biases, discovered in the social psychology literature since 1985.[154] The likelihood of multiple biases in individual behavior and cognition has both a qualitative and a quantitative impact on optimal policy.

A. Qualitative Effects

In this section we follow the analysis of Hanming Fang and Dan Silverman in showing that multiple biases working in the *same quantitative direction* have different implications for policy.[155] Suppose we were to design an optimal welfare policy that is paternalistic in the sense that it is best from the point of view of single mothers on welfare (rather than from the point of view of taxpayers). Suppose further, as is likely, that single welfare mothers have both excessive impatience and projection bias. In other words, they discount too heavily the delayed benefits of work—higher income and greater self-respect—relative to the immediate benefits of welfare, and they also overestimate the utility costs of work because they fail to predict their adaptation to working and, thus, its reduced irksomeness. These biases move in the same direction insofar as they reinforce the mother's desire to stay on welfare. Nevertheless, the logic of each bias is different and thus behavior will be differently affected. Present bias may be offset by inducing large and abrupt increases in the relative return to work through such policies as strict welfare time limits or immediate subsidization of work. This will circumvent the excessive discounting of future rewards. On the other hand, projection bias may be overcome by gently and slowly accommodating the transition to work through policies of gradual acquisition of human capital and exposure to work environments so that

153. *See generally* Gregory Besharov, *Second-Best Considerations in Correcting Cognitive Biases*, 71 S. ECON. J. 12 (2004).

154. *See* Joachim I. Krueger & David C. Funder, *Towards a Balanced Social Psychology: Causes, Consequences, and Cures for the Problem-Seeking Approach to Social Behavior and Cognition*, 27 BEHAV. & BRAIN SCI. 313, 317 tbl.1 (2004).

155. *See* Fang & Silverman, *supra* note 152, at 57.

the individual's preferences may more easily adapt to labor force participation. In order to determine which policy is best, the paternalist must have some idea of which bias is more important in the determination of behavior. Too much of one or the other policy can worsen the wellbeing of the single welfare mothers relative to their true, undistorted preferences. Abrupt policies might too quickly throw them off welfare when they are not adequately prepared in terms of human capital or acclamation to work. Gradual policies might keep them on welfare past the point where they would benefit from working.

At the present time, however, we do not know whether it will be possible to disentangle the *magnitude* of the biases from available data, even assuming that the two biases have been identified.[156] Moreover, it is not sufficient simply to try various policies and to endorse any policy that reduces welfare rolls. From a paternalist perspective, the goal is not simply to reduce welfare, but to reduce it optimally with respect to the single mothers' true preferences. Yet in the absence of knowledge of the complex interaction of the relevant biases, the appropriate policy prescriptions congruent with these preferences cannot be known.

B. Quantitative Effects

Behavior that seems suboptimal from the perspective of the measured bias may, in fact, be optimal when all of the biases are measured.[157] Even if it is suboptimal, it may not be suboptimal in the direction of the single measured bias. For example, even if individuals somewhat excessively discount the future costs of smoking, they may still smoke *too little*—in terms of their own long-run preferences—if they overestimate, perhaps due to availability bias, the health risks of smoking.[158] Identification of the former bias alone might lead the analyst to the conclusion that they smoke too much. Similarly, excessively impatient individuals may nevertheless save *too much* for retirement if they suffer from projection bias in assuming that future consumption tastes will be the same as at present, or if they do not accept the

156. *See id.* at 74 ("Moreover, we do not know yet, if the true generating process is a model with a combination of present and projection biases, whether we will be able to disentangle the magnitude of these biases from standard data.").

157. By "optimal," we mean here the most welfare-enhancing behavior *given* the existence of biases. This is second-best optimality. *See generally* Besharov, *supra* note 153.

158. On the possible overestimation of the health risks of smoking, see generally Fernando Antoñanzas et al., *Smoking Risks in Spain: Part I—Perception of Risks to the Smoker*, 21 J. RISK & UNCERTAINTY 161 (2000).

inevitability of death.[159] In the general case, the existence of multiple biases will make it difficult to determine the extent and direction of suboptimal behavior. To see this more clearly, consider the following examples.[160] Suppose an individual is subject to three biases: excessive impatience in the form of quasi-hyperbolic discounting, overconfidence about the favorable results of his actions, and ex post regret when he does not undertake sufficient effort to accomplish his goals. He is faced with a decision about a project that requires effort (costs) now and yields benefits in the immediate future. Excessive impatience tends to reduce his effort; overconfidence bias and regret bias tend to increase his effort.

Assume the paternalist knows the magnitude of just one bias, say, the overconfidence bias. He will then likely conclude that the individual's effort is above the optimum; and yet, due to the operation of the unobserved excessive impatience bias, it may actually be below the optimum. Paternalistic efforts to counter the overconfidence bias will thus exacerbate the suboptimal provision of effort.

Now assume that the paternalist knows the magnitude of *all* of the relevant biases. Yet the paternalist will not be able to determine the optimum level of effort toward attaining the agent's goal if he does not know both the value of the effort (imputed from the value of the goal) and the costs of effort. Since the biases are measured in different units, their impact on effort cannot be determined without knowledge of how much effort will be provided at various levels of the biases. Therefore effective debiasing would require a great deal of knowledge—so much so that a paternalist who possessed this would have to be near-omniscient.

Finally, assume that the paternalist knows the magnitude of all the biases, as well as the optimum level of effort, but not the individual's costs of correcting the separate biases. Presumably an individual who was aware of his biases would incur costs such that the usual condition of marginal cost equals marginal benefit is satisfied. In this correction equilibrium each bias may be treated differently—some may be reduced a good deal, some reduced just a little, and others may not change at all. Since the biases are interrelated, there may be second-best adjustments

159. On the effect of projection bias on saving, see Fang & Silverman, *supra* note 152, at 56. On the savings-and-consumption effects of the denial of death, see Kopczuk & Slemrod, *supra* note 134, at 4, ("Our model of death anxiety and the possible repression of information about mortality implies that people who are unaware of their denial will underconsume, acting as if their expected lifetime is longer than is accurate.").

160. These examples are taken from Besharov, *supra* note 153, at 18–19.

relative to those that cannot be cost-effectively changed. Whether these biases have already been optimally corrected will be hard to determine.

These examples make clear that even partial knowledge of a rather extensive nature is not sufficient to ensure welfare-enhancing paternalistic intervention. Clearly, to have all of the knowledge required, as even this simple model reveals, is out of the question.

X. ANTICIPATING UNRAVELING OF SELF-REGULATION AND THE SPREAD OF BIASES

A. Substitution Effects Between Internal and External Debiasing

Roughly speaking, there are two ways to solve a self-control problem: internally (through one's own efforts), or externally (through the efforts of third parties).[161] When the environment in which an individual makes his decisions is characterized by significant external control, the degree of self-control exercised will be lower. In other words, there is substitutability between external control and self-control.[162]

In the first instance, we can think of external control as purely social influence or pressure. For example, "individuals sometime criticize their friends or family members for eating unhealthy food or excessively watching TV."[163] Going one step beyond this, but still without legal coercion, "[s]ocial partners, groups and organizations may institute incentives, sanctions and rules that are designed to help individuals overcome temptations."[164] These factors are part of the local context that determines the degree of (counteractive) self-control exercised by individuals.

For example, students who were asked to take a 'diagnostic test of their reading skills' exercised varying levels of self-control depending on whether they were exposed to external pressure. When the test was characterized as *boring* and the students were not subjected to any external control or pressure, they exercised counteractive self-control by increasing their ex ante perception of the test's value.[165] Students that were asked to take the same test, characterized as *interesting*, on the

161. *See* Fishbach & Trope, *supra* note 148, at 256–59.
162. *See id.* at 260–61.
163. *Id.* at 256.
164. *Id.*
165. *Id.* at 260–61.

other hand, did not exercise this self-control and, therefore, had a relatively lower ex ante evaluation of the test. [166]

The exercise of counteractive self-control to convince oneself to take a boring test, however, appeared to break down when the students were subjected to external pressure. [167] Subjects that were asked to decide whether to take the boring test in the presence of the experimenter did not increase their evaluation of the test. [168] Rather, they decided to take the test as a result of the social pressure from experimenter-monitoring (external control) of the decision process.[169]

Thus, counteractive self-control and external control behaved as substitutes in influencing the subjects' decisions to take a boring test.[170] 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.[171]

We conclude from the above that self-control strategies and external control are interrelated. Individuals will adjust at their own margins depending on the exogenous context. Therefore, much of the experimental evidence showing self-control failures must be interpreted cautiously because looking at self-control alone does not give a complete picture. To be effective, policies designed to supplement deficient self-control with some form of paternalistic regulation must take account of the existing sources of external control. It may be that existing external pressures already maximally supplement the natural urge to exercise self-control—further external pressure might actually *decrease* the average person's predisposition to control herself or himself.

However, the policymaker's problem does not stop there, as policy itself might change important variables. While there do not seem to be direct studies of this problem, there is some suggestive research. Consider a study in which students were offered a "highly valuable" diagnostic test of their nighttime cognitive abilities.[172] The test would be

166. *Id.*

167. *Id.*

168. *Id.*

169. *Id.*

170. *Id.*

171. *Id.* at 261–63.

172. *See id.* at 263–66. Many students stay up late at night trying to study and so they are interested in this.

administered either at 9:00 PM or at the more inconvenient time of 1:00 AM. Half the students would be given a payment of twenty dollars to take the test while the other half would not receive any payment.[173] This is analogous to a government subsidy. (A policymaker who wanted to help people overcome short-run costs might subsidize the target activity.) The results are similar to those in the previously-mentioned studies. Those who were not offered the subsidy exercised counteractive self-control by increasing their evaluation of the test's importance.[174] Those who were offered the subsidy did not exercise self-control.[175] Thus there is another class of effects to consider. When external control in the form of payment is imposed, counteractive self-control decreases. Policy will itself change the level of self-control on which optimal policy depends.[176]

To summarize: policymakers who wish optimally to counteract deficient self-control need to know the amount of self-control that is being exercised under the status-quo. To know this, they must know the social-pressure context of the class of target decisions. As we have seen, most experiments are bad at replicating contexts in the wild.[177] If we let this problem pass, however, the policymaker still must know to what extent imposition of legal external controls will alter the status-quo of self-control as the context changes to one of more external control. We do not now have adequate information on the *degree* of substitutability between various types of external control and counteractive self-control to know whether particular policies will worsen or ameliorate the initial perceived deficiency of self-control.

173. *Id.*

174. *Id.*

175. *Id.*

176. Economists will no doubt prefer models in which there is a social optimum and an equilibrium level of a subsidy (tax) corresponding to an equilibrium level of self-control. Assuming such a model were applicable, a policymaker who knew the socially optimal subsidy could simply impose it, and the socially optimal level of self-control would be generated. The reality of policy is, however, far more messy. If a subsidy is imposed and people respond by reducing self-control, then arguments will be made for increasing the subsidy and expanding the degree of paternalistic intervention. Since no one is likely to know the optimal level of self-control, the process might simply continue until the subsidy replaced self-control entirely.

177. *See supra* Part VII.B.1.

B. Generalized Reduction of Self-Regulation

Many psychologists believe that the capacity for internal control is a scarce resource subject to depletion.[178] If individuals have previously exercised self-control, immediate subsequent efforts at self-control will be less successful.[179] Thus when individuals exercised some self-regulatory effort in an initial task, they were then more likely to "spend money impulsively . . . [,] show higher levels of aggressive responding . . . [,] drink more alcohol even when anticipating a driving test . . . [,] [and] perform inappropriate or uncontrolled sexual behaviours . . . [,]'' as well as engage in a wide-range of other low self-regulation activities.[180] All of this is consistent with a short-run fixed supply of self-regulation.

If, as we have previously argued, external control substitutes for self-control, and if self-control is a limited resource, then, plausibly, an increase of external control might release some self-control capacity for other tasks from which it had been missing. In other words, any loss in self-control occasioned by the adoption of paternalist policies in one area of life might be offset by increases in self-control for other areas of life.

In the longer run, however, lack of exercise of self-control capacity leads to a decline of that capacity.[181] In other words, although the supply of self-control is fixed in the short run, it is not in the long run.[182] The capacity for self-control can be augmented in the long run by its exercise in the short run. To put the issue in metaphorical terms, self-control is more like a fund in the short-run, but more like a muscle in the long run. In the short run, you can run out of self-control; in the long run, exercise can augment your self-control.

Consider the following representative experiment. Researchers assessed the motivation of people to avoid the expression or appearance of prejudice toward homosexuals and obese people.[183] Consistent with previous findings, some people were highly motivated and others were

178. *See, e.g.*, Roy F. Baumeister et al., *Self-Regulation and Personality: How Interventions Increase Regulatory Success, and How Depletion Moderates the Effects of Traits on Behavior*, 74 J. PERSONALITY 1773, 1773–76 (2006).

179. *Id.* at 1776.

180. *Id.* (citations omitted).

181. *Id.* at 1779–86.

182. *Id.*

183. *See, e.g.*, Matthew T. Gailliot et al., *Increasing Self-Regulatory Strength Can Reduce the Depleting Effects of Suppressing Stereotypes*, 33 PERSONALITY & SOC. PSYCHOL. BULL. 281, 283–86 (2007).

minimally motivated to avoid prejudice.[184] Participants were then asked to write about a day in the life of a hypothetical homosexual or obese person without resorting to stereotypes.[185] In effect, they had to use, to a greater or lesser extent, self-regulatory capacity to suppress the stereotypes. Afterwards, in another task, the same individuals were asked to solve anagrams.[186] This required the further exercise of self-regulation. In general, people performed worse on the second task.[187] The worst performance, however, was from those who had displayed, in the first task, low self-regulatory traits in avoiding prejudice.[188] For them the cost of the initial suppression task was high and, in the short run, greatly depleted their self-regulatory capacity.

To capture longer-run effects, participants were asked to practice self-regulatory activity, unrelated to stereotype suppression, for two weeks.[189] Then participants were retested to determine the degree to which the primary self-regulatory activity—suppressing stereotypes—depleted capacity with respect to subsequent self-regulation.[190] The important finding is that the two weeks of *unrelated* exercise of self-control increased the performance on the second task.[191] Thus, practice in the short run increased self-regulatory capacity in the longer run. Furthermore, this increase was seen only in the individuals who had a low propensity to avoid prejudice, that is, only in those who had an initially high cost of suppressing stereotypes.[192]

The first conclusion we draw is that policies that decrease the exercise of self-regulation in the short run will decrease the amount of self-regulatory capacity in the longer run. Secondly, this decrease will manifest itself in areas unrelated to the initial decrease in self-regulation. For example, lesser (or greater) self-management in financial affairs can affect self-regulation in the same direction in the areas of diet, smoking, and alcohol consumption.[193] The third conclusion is that individuals with initial high costs of self-regulation benefit more from exercise of short-

184. *Id.*
185. *Id.*
186. *Id.*
187. *Id.* at 286.
188. *Id.*
189. *Id.* at 286–88.
190. *Id.*
191. *Id.*
192. *Id.*
193. *See generally* Megan Oaten & Ken Cheng, *Improvements in Self-control from Financial Monitoring*, 28 J. ECON. PSYCHOL. 487 (2007).

run self-control than others. Or, to put things negatively, those who have a high cost of self-regulation have the most to lose from the short-run substitution of external for internal control.

How do these conclusions affect the paternalist policymaker's problem? He must recognize that supplementing self-control with external control in a particular area will, in the longer run, lead to the decrease in self-regulatory capacity and the spread of deficient self-control to other, unrelated areas. This will reduce or perhaps negate completely the benefits of a paternalistic intervention. But it will not do this uniformly. The effect will be greater the larger the initial costs of self-control. All of these effects are difficult to account for, because the contextual nature of self-regulation means these effects are contingent on local facts. As we saw in the last section, the degree to which greater external regulation will crowd out short-run self-control will depend on the relative efficacy of each. Now we see that the degree to which reduced short-run self-control will result in lower long-run self-control, and the areas to which it will spread will depend on initial self-control propensities in particular areas. Once again, the local knowledge issues threaten the facile policy use of the generalizations from behavioral economics.

XI. ACCOUNTING FOR HETEROGENEITY: THE ONE-SIZE-FITS-ALL PROBLEM

Knowing that a bias exists is not enough. Knowing the extent of bias for a particular individual, or for the typical individual, is also not enough. For the paternalist to construct effective policies, the paternalist must also take into account the heterogeneity of individuals in their decision-making biases.

A. Problems of Over-Inclusion and Under-Inclusion

Most, if not all, proposed policies have a "one-size-fits-all" flavor, in that they cannot be targeted at specific individuals. As a result, most policies will tend to create problems of both under- and over-inclusion, meaning that some people whose behavior needs correction will not be affected enough, while other people whose behavior requires less change (or no change at all) will be affected too much. A fat tax, for instance, would apply to all buyers of food. Some overeaters will continue to eat too much because the fat tax is insufficiently large (or because they are indifferent to the tax), while some non-overeaters will be induced to reduce their consumption unnecessarily, with a resulting reduction in

satisfaction. Whether the gains from those helped exceed the losses to those harmed by a policy depends crucially on the distribution of the extent of bias across the affected population—which means the paternalist policymaker needs extensive information about that distribution. (It will also depend crucially on making interpersonal comparisons of utility, a problematic matter in and of itself.) Policymakers could, of course, try to create special exemptions (total or partial) for those deemed not to require special assistance in correcting their biases. But a finely tuned policy of this nature would require a great deal of information in order to identify which individuals to grant exceptions (and to what extent).

There is abundant evidence that both behavioral and cognitive biases are not uniform.[194] They are distributed in the population along such parameters as performance on the Scholastic Aptitude Test (possibly a measure of general cognitive ability),[195] cognitive mindsets or dispositions,[196] cultural differences,[197] and gender differences.[198] Affective changes within a single individual as well as, possibly, developmental changes can also affect the existence or degree of biases.[199] All of this will complicate the determination of optimal policy where, as we see below, policy cannot be tailored according to the individual's characteristics. Additionally, and perhaps most importantly for policy prescriptions, individuals may differ substantially in their behavior from situation to situation.[200] This implies that measured biases in one area will be inaccurate if applied to other areas, and thus optimal policy will be different according to, for example, whether we are dealing with junk food consumption or savings behavior.[201]

As we have seen previously, the quasi-hyperbolic discounting literature claims that individuals have a long-run rate of time discount

194. *See, e.g.*, Gregory Mitchell, *Why Law and Economics' Perfect Rationality Assumption Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence*, 91 GEO. L.J. 67 (2002) (citing at least one hundred studies).

195. *Id.* at 94–95.

196. *Id.*

197. *Id.* at 147–56.

198. *Id.* at 140–46.

199. *Id.* at 156–60.

200. *Id.* at 105–19.

201. For example, the rate of time discount applied to choices in different areas may vary. *See* Frederick, Loewenstein & O'Donoghue, *supra* note 22, at 394 ("Since different motives may be invoked to different degrees by different situations (*and by different descriptions of the same situation*), developing descriptively accurate models of intertemporal choice will not be easy.")(emphasis added).

corresponding to their true intertemporal preferences as well as an excessive rate corresponding to their lack of willpower. Those with self-control problems will give the future negative consequences of their actions less weight than they should. In particular, individuals may consume goods with large current benefits and significant long-term health costs because they lack the power to resist temptation.[202]

The difficulties posed by heterogeneity of individuals are best illustrated by the policy of sin taxes. Following O'Donoghue and Rabin, let us call the generic sin good to be taxed "potato chips."[203]

In a world with no costs of determining or collecting taxes, the first best optimum would be for the state to impose an individually calibrated tax on each individual corresponding to his degree of excessive impatience and the negative health consequences of potato-chip consumption. Then, the benefits to the present self of potato chips would be reduced by the negative consequences to future selves now made present by the tax. But obviously, this is not a practical suggestion. The paternalist is really faced with the necessity of determining a single or uniform tax rate that will apply to everyone regardless of his particular degree of excessive impatience. The tax will be too high for some, too low for others, and for a few just right.

The problem that is faced by the paternalist is to find the uniform tax rate that will minimize the cost of "errors" committed by the consumer.[204] The first error is that of *over*-consumption of potato chips, and the second error is the *under*-consumption of potato chips. Not every reduction in potato chip consumption by those who are consuming too much in the no-tax status quo is a benefit, because some may decrease their consumption too much. And some, without self-control problems, may be consuming just the right amount under the status quo. Therefore, the benefits of reducing potato chip consumption towards the optimum must be balanced against the costs of reducing consumption too much.

What must the paternalist policy maker know in order to determine whether a proposed tax rate will enhance or reduce welfare relative to the no-tax baseline? In the general case, he must know the *distribution* (the population heterogeneity) of the degree of self-control bias.[205] Some individuals will have greater self-control problems than others, and some

202. *Cf.* O'Donoghue & Rabin, *Optimal Sin Taxes*, *supra* note 3, at 1826.

203. *See id.*

204. For concreteness and precision, we follow the basic structure of the model developed in O'Donoghue & Rabin, *Optimal Sin Taxes*, *supra* note 3.

205. *See id.* at 1841.

will have no self-control problem as we conceive of it here. In addition, he must know something about the heterogeneity in people's tastes for potato chips and their susceptibility to adverse health consequences.[206] Thus, the distribution of immediate consumption benefits and future health costs must be known. Furthermore, the paternalist must know the elasticity or responsiveness of consumption at different tax rates in order to determine how much a given increment in tax will reduce consumption.[207] It is particularly important to know whether the *degree* of self-control problem is correlated with responsiveness because, if it is, a given tax will reduce consumption by different amounts by those with greater or lesser control problems. All of these factors will affect both the optimal tax rate *and our ability to know whether we have improved matters overall.* The problem is that we do not have, and are not likely to get in the near future, reliable data on these parameters.[208] In addition, there will no doubt be different relevant distributions for different kinds of sin goods. Cigarettes, fatty hamburgers, transfat french fries, hard liquor, lack of exercise, sugary deserts, and refined carbohydrates are different areas with different temptations and consequences.

In actual policymaking, the likely result of these complications is that they will be ignored. The paternalist will, in practice, be satisfied if potato chip consumption simply falls with no thought of the costs. Thus his preferences will supplant those of the individuals. Once again, the new paternalism in theory will be more like the old paternalism in practice.

B. Heterogeneity on Multiple Dimensions

As we have seen, when individuals differ with respect to a single bias, the policymaker's task is complicated by the need to calibrate the policy—say, a fat tax—to the population distribution of that bias and not to the single or average case.[209] However, as we have also seen, people exhibit more than one bias at a time. Each bias is itself not uniform across individuals. The policymaker's problem now becomes the calibration of policy to the *distributions* of *multiple*, possibly conflicting or reinforcing, biases. The optimizing mathematics of this situation is no

206. *Id.*

207. *Id.*

208. It is clear from their discussion that O'Donoghue and Rabin are simply making "back of envelope" calculations in their own example with no pretense of empirical accuracy. *See id.* at 1836–39.

209. *See id.*

doubt complex. The insurmountable character of the problem becomes apparent when we recognize that *individuals will exhibit heterogeneity along every dimension discussed thus far in this article*. Given the dearth of research on this topic, we offer only a partial list of the relevant ways in which individuals differ.

1. Fraction of individuals exhibiting a type of bias

Even if most individuals are subject to some sort of bias, not every individual will be subject to the very same biases.[210] Some have greater problems with weakness of will; others are most susceptible to making rash choices in hot states; yet others are most likely to fall prey to framing effects. The paternalist policy designer needs to know what fraction of the population falls into each category of bias. A larger fraction will tend to justify more, and more extensive, interventions, while a smaller fraction will justify fewer, and less extensive, interventions.

2. Extent of bias

As discussed earlier, optimal paternalist policy depends on knowledge of the extent of a bias, not merely its existence. *Yet the extent of bias will differ across individuals*. Among those subject to emotional (hot-state) decision-making, some will be more rash than others and have more to regret later. Among people with willpower problems, some people have bigger willpower problems than others. Among people with impatience problems, some will have greater impatience and others will have less.

3. Extent of self-debiasing

Individuals who are aware of their own bias problems will often try to correct them. But by their very nature, self-debiasing efforts are idiosyncratic. People will differ in their self-debiasing efforts by (a) the methods chosen, (b) the areas of life in which they have attempted to debias, (c) the extent of interdependence of their debiasing methods across areas, and (d) their degree of success—or even over-success, in the case of individuals whose resolutions and commitments turn into self-denying compulsions.

210. *See* Krueger & Funder, *supra* note 154, at 317 tbl.1.

4. Degree of responsiveness to corrective measures

People will differ in how much they respond to externally imposed debiasing policies. Some biases may be so strong or resistant to correction that costs—including externally imposed ones—are simply ignored. For example, a severe overeating problem could result from a strong propensity to underweight future costs relative to present benefits. A strongly present-biased person might care as little about future wealth as future health, whereas a mildly present-biased person might care a great deal about future wealth. If so, then a fat tax would have little effect on the former and a large effect on the latter. Effects like this have been observed with respect to existing sin taxes; for example, it turns out that moderate drinkers are more responsive to changes in price than are heavy drinkers.[211]

5. Susceptibility of self-debiasing to unraveling

Given that self-control and external control can act as substitutes, the extent of their substitutability will matter for policy. But the extent of substitutability will also differ across individuals. Some people will substantially reduce their self-control efforts in response to paternalist policy, while others may reduce their self-control little or not at all.

XII. CONCLUSIONS: THE ROAD BACK TO OLD PATERNALISM

Let us make a short recapitulation of the many forms of knowledge that a paternalist policymaker must possess in order for his policies to have any reasonable expectation of improving welfare. First, the paternalist must know individuals' "true" underlying preferences—which, by the paternalist's own hypothesis, are not (simply) revealed by choices. In doing so, he must choose between different and conflicting preference sets that seem to motivate individual behavior under different circumstances, without any firm theoretical means of doing so. Second, the paternalist must discover the extent of any given bias, understanding that any given bias will differ from time to time, place to place, and situation to situation—even for a single individual. Third, the paternalist must possess extensive knowledge of the self-debiasing measures adopted by individuals. Such measures come in a wide variety of forms and often depend on contextual features of the environment. Fourth, the

211. Brent D. Mast, Bruce L. Benson & David W. Rasmussen, *Beer Taxation and Alcohol-Related Traffic Fatalities*, 66 S. ECON. J. 214, 217 (1999).

paternalist must account for the interdependence of biases. This means that even comprehensive knowledge of a single bias is not sufficient to justify paternalist correction of that bias; the paternalist must understand the complex interaction of multiple biases. Fifth, the paternalist must anticipate and account for how paternalist policies may reduce the extent of self-regulation, both in the targeted field of activity and others as well. And sixth, the paternalist must possess all of the above kinds of knowledge not merely at the individual level, but at the level of the whole population. Knowledge of averages or general tendencies is not sufficient, as any given policy will affect people in different and sometimes offsetting ways.

One obvious defense of the new paternalist project is to say we simply need to collect more information. This, in itself, constitutes a major concession; it means recent proposals for paternalist interventions should at least be put on hold until superior information becomes available. But more importantly, this defense fails because much of the necessary knowledge is unavailable to a paternalist planner *in principle*. The relevant information about the extent of real-world biases is necessarily *local* in character; that is, it depends on particular characteristics of time and place. It changes from moment to moment and situation to situation. It differs substantially across individuals. It is affected by multifarious forms of self-regulation. It generally cannot be collected in a laboratory setting, because decision biases "in the wild" are what matter for policy. But in the wild, as opposed to the lab, there usually does not exist a means of holding other factors constant in order to "fix" the individual's true preferences and thus to measure deviations from them.

Moreover, much of the necessary information is *tacit*, meaning that it cannot be communicated easily. An individual might have great difficulty explaining what things are most tempting to him even if he wanted to. Some forms of self-correction are unconscious, occurring in ways that the individual is not even aware of. And no amount of data collection can overcome the theoretical problem of selecting among competing preference sets held by a single individual. Even if it is granted that an individual has "true" preferences, the paternalists have not yet enunciated a clear means of determining which preferences are true. The true preferences, by their very nature, exist only within an individual's brain and, as the new paternalists themselves insist, they are not straightforwardly revealed by choice.

Another defense of the new paternalist project is to deny that so much knowledge is really needed. According to this defense, all policymakers really need is a knowledge of averages or general tendencies. They might not be able to craft perfectly optimal policies, but they can make marginal changes that will improve welfare relative to the status quo. This defense is simply mistaken, largely because of the effects of heterogeneity. When a policy will produce positive effects for some and negative effects for others, only a knowledge of the distribution of such effects is sufficient to make a prima facie case that an intervention is welfare-improving. Even knowing that the average or typical person is in need of paternalistic assistance is not sufficient because (a) the average or typical person could be less responsive to corrective measures than others who do not need the assistance or who need it less; or (b) the average or typical person might respond in counterproductive ways, such as reducing self-corrective efforts.

In any case, both defenses just offered rely on an excessively optimistic conception of the political process. They imagine careful and comprehensive investigation by intelligent, well-meaning, and motivated political actors. The reality would assuredly be much different. Faced with daunting, and often insurmountable, barriers to accessing and processing all the information they need, politicians and bureaucrats will more likely rely on rules of thumb. Lacking information about true preferences, they will tend to appeal to their own preferences or to socially approved preferences.

For instance, what would be considered evidence of a real-world anti-obesity measure (like a fat tax) having been effective? Keep in mind that behavioral economics emphatically does not indicate that obesity is necessarily an irrational decision. An honest and accurate measure of an anti-obesity measure's efficacy would have to measure (somehow) both the gains to people who are nudged closer to their true preferences *and* the losses to people who are nudged further away—including people who are not obese but who are motivated to change their behavior anyway, as well as to people who really are obese but whose true underlying preferences justify their condition. We are not sure how the government would even begin to collect such information. But no matter, because in the real world of politics we suspect that only falling rates of obesity will suffice. "Eating right" is the socially approved preference.

And thus the new paternalism transforms, in practice, into the old. In principle, we can embrace the idea of making people better off according to their own true preferences. That goal cannot be made operational in

practice without access to information that policymakers do not, will not, and often cannot possess. Yet policymakers have to make policy on the basis of *something*, and so they will appeal to their own preferences, the preferences of self-appointed experts, or the (alleged) preferences of the public at large. They cannot implement people's "true" preferences, but they can implement what they believe are the "right" ones, and the new paternalist paradigm will provide the intellectual cover to do so.