Policy Analysis

No. 563 February 22, 2006

Routing

Against the New Paternalism
Internalities and the Economics of Self-Control

by Glen Whitman

Executive Summary

Economists have long argued that government intervention makes most sense in situations that involve externalities. Externalities are costs or benefits that spill over onto third parties. When individuals bear the full costs and receive the full benefits of their own actions, the justification for government involvement is much weaker. But a new generation of economists contends that paternalistic intervention can be justified to correct problems of self-control. If people don't fully consider the costs their choices impose on their own future selves, the theory goes, those choices impose within-person externalities dubbed "internalities." The internalities approach provides a novel argument in favor of paternalistic government policies such as sin taxes (including fat taxes), marketing restrictions, mandatory savings plans, and so on.

The theory of internalities is explicitly modeled on the theory of externalities. However, the former stands about where the latter stood in 1960, just prior to Ronald Coase's seminal work on the subject. Exposing internality theory to Coasean insights reveals serious flaws. Specifically, internality theory in its current form unjustifiably "takes sides" when it chooses to favor some personal interests over others. Furthermore, it ignores the possibility of within-person bargaining and other private solutions to self-control problems. Finally, it gives insufficient attention to the possibility of government failure. Taking those objections into account severely damages the case for paternalistic government intervention to address problems of self-control.

Glen Whitman is associate professor of economics at California State University, Northridge.

CATO INSTITUTE

Because the consumer bears most of the cost of eating Twinkies, he is well positioned to decide whether their perceived benefits outweigh their harms.

Introduction: How Is a Twinkie like a Smokestack?

How is a Twinkie like a smokestack? Set aside the vaguely similar shape and think about the harms they create. The harm from a Twinkie falls primarily on the consumer, in the form of worse health and larger hips. Because the consumer bears most of the cost of eating Twinkies, he is well positioned to decide whether their perceived benefits outweigh their harms. A smokestack, by contrast, affects more than the factory and the consumers of its products—specifically, it harms everyone in the surrounding community who has to breathe the air. The firm's management might not, in its pursuit of profit, take into account all of the costs associated with its polluting activity.

The difference in who is harmed—the decisionmaker versus someone else—might seem an important distinction. And indeed, economists have generally held that the case for government intervention is strongest when third parties are involved—that is, when there are *externalities*. Externalities are costs or benefits of an activity that spill over onto people not involved in the activity (typical examples include people who breathe polluted air or neighbors disturbed by loud music). Self-regarding activities, on the other hand, can safely be engaged in by free individuals. But a growing literature in economics, as well as the popular press, argues that Twinkies and smokestacks are more similar than they appear. Although your choice to eat Twinkies or smoke cigarettes or skip exercising today doesn't generally harm anyone else, it does harm your *future* self. If we think of a person as consisting of multiple selves—the present self who wishes to indulge in transient pleasures versus the future self who wishes to be healthy—then arguably the present self's choices can force externalities on the future self. Those within-person externalities have been dubbed "internalities." And just as we might impose a pollution tax on a factory to control the externality problem, we might impose a sin tax on items like cigarettes, alcohol, and fatty foods to control the internality problem.

The concept of internalities, although not yet a part of mainstream economics, is gaining attention. It is one among many novel economic models recently deployed by a new generation of paternalists. Paternalistic arguments advocate forcing or manipulating individuals to change their behavior *for their own good*, as distinct from the good of others. At one time paternalists argued that adults, like children, don't really know what's best for them. Some preferences, they argued, such as those for unhealthy food or casual sex, are just wrong. But such arguments hold little sway in a free society, where most people believe they should be able to pursue their own values and preferences even if others don't share them. So the "new" paternalists have wisely chosen not to question people's preferences directly; instead, they argue that internalities (and other sources of error in decisionmaking) can lead people to make decisions that are unwise *even according to their own values and preferences*.

In short, the old paternalism said, "We know what's best for you, and we'll make you do it." The new paternalism says, "*You* know what's best for you, and we'll make you do it."[1]

The internalities approach is clever. Even the staunchest skeptics of government intervention will usually concede that government is needed to prevent people from harming each other. By treating the individual as a multiplicity of selves, the new paternalism invites policy analysts to import the theory of externalities into the realm of individual choice.

Of course, thinking of a person as having multiple selves is a controversial philosophical position, and we might be tempted to reject it outright. We could object that "multiple selves" is, at best, just a metaphor. But here, I wish to make an *immanent* critique. I will take the idea of multiple selves as given—for argument's sake—and then argue that the analysis of internalities is seriously incomplete.[2] In "translating" the concept of externalities, internality theorists have drawn on economic theo-

ry that is at least 40 years out of date. In his famed 1960 article "The Problem of Social Cost," Nobel Prize–winning economist Ronald Coase started a revolution in the economic analysis of externalities.[3] His approach showed that externalities need not always produce inefficient outcomes, because both institutions (such as private property rights) and market exchanges can deal with them. The application of Coase's ideas to internality theory casts serious doubt on the new paternalism.

Internalities without Coase

Prior to the Coasean revolution, externality theory followed the analysis of A. C. Pigou.[4] The Pigovian argument is straightforward:

- In the presence of externalities (sometimes called "spillover" effects), private decisionmakers do not face all the costs of their choices.
- As a result, they will do too much of the activity in question. Factories will produce additional goods even when the added revenue doesn't cover the added cost, for instance, because part of the cost is foisted onto the air-breathing public. That is inefficient.
- An appropriately chosen tax on the activity in question can correct the inefficiency. If the factory has to pay a tax for each unit of output, or for each cubic foot of soot it pumps into the air, it will adjust its activity accordingly.

Internality theorists take those three claims and apply them, with little change, to choices *within* the individual. They say that individuals have a systematic tendency to pay too little attention to costs and benefits of decisions to their *future* selves. As a result, they will engage in excessive amounts of certain activities, such as smoking, eating, and drinking. Taxes on those activities will reduce them, thus making people better off. To illustrate:

- Jonathan Gruber and Botond Koszegi contend that cigarette smoking produces negative internalities, argue that government policies "should depend not only on the externalities that smokers impose on others but also on the 'internalities' that smokers impose on themselves," and calculate that "there are sizable optimal 'internality' taxes on the order of $1 per pack or more."[5]
- Ted O'Donoghue and Matthew Rabin discuss "optimal sin taxes" designed to correct self-control problems, using the specific example of overeating.[6] In their not-yet-published extension of that article, they note that "since people with self-control problems impose negative externalities on their future selves—dubbed 'negative internalities' . . . the role that sin taxes play in our analysis is much like a Pigovian tax to correct negative externalities."[7]
- Other authors, such as Colin Camerer et al.[8] and Richard Thaler and Cass Sunstein,[9] have offered more tentative policy prescriptions, preferring to weaken the prima facie case against paternalism. Recent research in behavioral economics, they say, "potentially broadens the scope of situations in which paternalistic policies could usefully be developed."[10] Still, they invoke the term "internalities" along with its hefty market-failure baggage: "When consumers make errors, it is as if they are imposing externalities on themselves because the decisions they make . . . do not accurately reflect the benefits they derive."[11]

All of these authors use old-fashioned Pigovian externality arguments, with no reference to Coase or his insights. Yet Coase threw a monkey wrench into Pigou's works long ago.

The Reciprocal Nature of Internalities

The question is commonly thought of as one in which A inflicts harm on B and

By treating the individual as a multiplicity of selves, the new paternalism invites policy analysts to import the theory of externalities into the realm of individual choice.

> what has to be decided is, How should we restrain A? But this is wrong. We are dealing with a problem of a reciprocal nature. To avoid harm to B would be to inflict harm on A. The real question that has to be decided is, Should A be allowed to harm B or should B be allowed to harm A? The problem is to avoid the more serious harm.[12]

Imagine that the individual consists of more than one self. One common approach says that the individual contains many different selves, one present-oriented self for each moment in time and a single future-oriented self.[13] Another approach imagines just two selves, a single present-oriented self and a single future-oriented self. When it eases explanation, I will use the simpler model.

Coase's first major insight concerning externalities was that they are *reciprocal* in nature. In other words, an externality is not simply a harm imposed by one person on another. Rather, an externality arises because two (or more) people have conflicting interests, and championing the interests of one party means denying the interests of the other.

Take, for example, the case of a residential neighborhood near an airport. The incoming and outgoing flights create a noise nuisance for neighboring homes; this is a negative externality. Allowing the airport to continue operation causes harm to the residents. But it also works the other way around: restricting flights to give residents more peace and quiet does harm to the airport (and its customers). This point becomes especially clear when we note that the residents often move into the area around an airport voluntarily, *after* the airport is already in operation. Apparently, such residents think the benefits of living near the airport compensate for the noise.

The application of Coase's ideas to internality theory casts serious doubt on the new paternalism.

But Coase's argument does not depend on who was there first. The point is simply that *harm is a two-way street*. The existence of an externality tells us that a tradeoff exists between some people's interests and others'. It does *not* tell us how the tradeoff should be made. In some cases, it might make sense to shut down an airport (or restrict its flights) because the cost of the noise exceeds the benefit. In other cases, it will be better to allow the airport to operate unimpeded and let the residents adjust—by moving elsewhere, soundproofing their homes, or putting up with the noise in exchange for lower housing costs.

Internality theorists observe that the short-run self can take actions, such as smoking or overeating, that will harm the interests of the long-run self. They predict the short-run self will therefore engage in too much of the harmful activities at the expense of the long-run self. This approach is fundamentally Pigovian. It regards one actor (the short-run self) as the sole cause of the harm, and the other actor or actors (the long-run self or all future short-run selves) as the passive victims of that harm. As Coase observed in the quotation above, that analysis is one-sided. True, allowing the present self to smoke or overeat means harming the future self. But, by the same token, preventing smoking or overeating on behalf of the future self means harming the present self.

To take the notion of multiple selves seriously, the analyst must consider both sets of interests or preferences. We should not simply assume that the long-run self's interests somehow supersede those of the short-run self, any more than we should assume the residents' interests supersede those of the airport. Thus, adopting policies *solely* on grounds that they advance the interests of the long-run self would be inappropriate.

As Coase observes, the real problem is to avoid the more serious harm. But nothing about the situation, certainly not the fact that the short-run self may impose harm on the long-run self, shows that the long-run harm exceeds the short-run harm. This becomes apparent if we consider that the long-run self can harm the short-run self by adopting self-control devices—such as flushing cigarettes down the toilet, refusing to allow ice cream in the house, checking into a clinic, and so on. The future long-run self may also impose the cost of guilt on the present self.[14] Such actions help the long-run

self at the expense of the short-run self. Given the reciprocal nature of the problem, and having no further information, we could just as easily conclude that the long-run self imposes internalities on the short-run self that require correction. Perhaps we should tax weight-loss clinics.

Going a step further, we could observe that future-oriented selves sometimes induce behaviors that, at least to outside observers, appear excessive. In contrast to the obese and the profligate, whose short-run selves constantly trump their long-run selves, we might point to the misers, workaholics, and anorexics for whom the reverse appears to be true.[15] Even among "normal" individuals, studies show that excessive self-control efforts can lead to underconsumption of desirable things.[16] Perhaps we should subsidize Krispy Kreme.

Or, following Coase, we could recognize that harm goes both ways. The existence of an interactive effect does not, in itself, tell us that an inefficiency exists; nor does it tell us whether the inefficiency, if any, results from too much or too little of an activity.

Time Inconsistency and Multiple Selves

Where does this notion of "multiple selves" come from? You're just one person—right? It might seem that way, but internality theorists say your actions betray you. People often make choices that reflect internal conflict, or tension between different sets of preferences. Specifically, people display *time inconsistency*, which (in simple terms) means a conflict between your preferences tomorrow and your preferences today.

Suppose, for example, that you're offered a choice between $100 to be received 100 days from now, or $110 to be received 101 days from now. Many people will choose the larger amount of money. But now take that same choice and move it forward 100 days, so that you're choosing between $100 today and $110 tomorrow. Given that choice, many will choose the smaller amount—including people who chose the larger amount for the more distant choice.[17] This phenomenon, known as preference reversal, can lead to inconsistent behavior, such as making promises for the future ("I promise to start my diet Monday") and then breaking them when the date arrives ("No, I guess I'll start my diet *next* Monday"). And lest it seem that those are just idle promises, people will even *limit their own future options* to make promise breaking more difficult—for instance, by emptying their pantries of tempting snacks.

Time inconsistency, according to internality theorists, shows the existence of competing interests within the self—or, more dramatically, competing selves. And this is not just a run-of-the-mill tradeoff, like the tradeoff between watching TV and going to the movies. Rather, time inconsistency means that the rate of tradeoff itself changes systematically over time. It reveals a kind of schizophrenia in the individual, albeit a schizophrenia present even in the most normal people. The impatient short-run self places more weight on immediate gratification, whereas the long-run self places greater weight on delayed gratification. And then, the argument goes, the short-run self takes advantage of its control of the body to foist harms on the helpless long-run self.

Although time inconsistency does reveal a kind of internal conflict, it tells us nothing about how to resolve the conflict. Look again at the choice between $100 one day and $110 a day later. If someone chooses the smaller amount when the choice is near but the larger amount when the choice is distant, we could "correct" him by manipulating him to choose the larger amount always. That would make his choices consistent—and this is, in essence, what the internality theorists think we should do. But we could also "correct" him by making him choose the smaller amount always. That, too, would make his choices consistent. So there is more than one way to "fix" a time inconsistency, and internality theory tells us nothing about which fix to use.[18] Once again, we observe the reciprocal nature of internalities: assisting one set of preferences means harming the other.

We should not simply assume that the long-run self's interests somehow supersede those of the short-run self.

In contrast to the obese and the profligate, whose short-run selves constantly trump their long-run selves, we might point to the misers, workaholics, and anorexics.

The Least-Cost Avoider Principle

> Suppose that those who suffer the damage could avoid it by moving to other locations or by taking various precautions which would cost them [less than the polluter's avoidance cost]. Then there would be a gain in the value of production . . . if the factory continued to emit its smoke and those now in the district moved elsewhere or made other adjustments to avoid the damage.[19]

Typically, there exists more than one means of averting a harm. The object is to induce action by the *least-cost avoider* of harm.

Consider a classic externality story: a cement factory spews dust on the residences that surround it.[20] A simple Pigovian analysis says the factory creates harm to the residents and, therefore, ought to be taxed for the dust it creates. A Coasean analysis points out that the dust nuisance might be avoided or reduced in more than one way. The factory could shut down or reduce its production. Alternatively, the residents could move away, not move there in the first place, or act to mitigate the dust's impact (by not hanging their washing outdoors, for instance). Which course of action ought to be taken? Nothing in the description of the situation gives us the answer. In some situations, it would be cheaper for the factory to reduce its output (or shut down) than for the residents to move (or reduce their exposure in some other way). In other situations, it would be cheaper for the residents to change their behavior. In the latter situation, where the residents are the least-cost avoiders, a tax on the factory would not improve the situation. The tax would tend to reduce the factory's production, even though the value of the lost production would exceed the cost to residents of averting the same harm.[21]

Analogously, the harm resulting from an internality might be avoided in more than one way. The short-run self could reduce its Twinkie consumption, eat a Twinkie Lite instead, or have it with a Diet Coke instead of a Coke. Alternatively, the long-run self could adopt measures designed to reduce the Twinkie's future effects. It could, for instance, commit to exercising more often (or more vigorously) by joining a gym or making agreements with workout partners. Or the long-run self might resign itself to taking heart medications. Which route is most efficient depends on the subjective cost of the different options. If the future-oriented self were the least-cost avoider, a Twinkie tax would not improve matters. It would induce the present self to eat fewer Twinkies, even though the future self could have avoided or reduced the harm at a lower cost.

Returning to the cement factory example, there is a third outcome that might, depending on the parameters, prove efficient: doing nothing. If the value of the factory's output (which would be lost if the factory shut down) is greater than the damage done, and any avoidance measures by the factory or residents would impose costs greater than the damage avoided, then it makes sense to create dust with no avoidance measures at all. Analogously, if the value of the Twinkie to the short-run self is greater than the damage done to the long-run self, and avoidance measures by either self involve costs higher than the damage they avoid, it makes sense to eat the Twinkies without countermeasures.

Property Rights and Exchange

> It is always possible to modify by transactions on the market the initial legal delimitation of rights. And, of course, if such market transactions are costless, such a rearrangement of rights will always take place if it would lead to an increase in the value of production.[22]

Assuming the harm exceeds the avoidance cost, what can induce the least-cost avoider to take the appropriate action? Coase sug-

gests the possibility of transactions between the parties. If the cement factory is the least-cost avoider, but the factory is allowed by law to pollute, the residents can pay the factory to shut down or cut back production. More broadly, Coase's point is that people can find creative ways to negotiate with each other to exploit opportunities for gain. Could a similar solution apply to internalities?

I contend that the answer is yes. Just as different people can make deals, different selves within a person can make deals. I will call such deals *intrapersonal bargains*. Those bargains can be struck in at least three different ways.

Collusion for Mutual Gain

This sort of intrapersonal bargaining, first explained in detail by psychologist George Ainslie,[23] relies on a specialized form of cooperation among a succession of present selves. Ainslie observes that even your impatient present self cares *to some degree* about your future selves—just not enough to make the preferences of present and future selves perfectly consistent. Your present self would actually like to see greater self-control, because that would benefit all future selves, which the present self cares about, too. The self-control problem arises because the present self would like to make a special exception for itself. In choosing between starting a diet today and starting a diet tomorrow, the present self prefers the latter. But the same goes for tomorrow's present self once tomorrow arrives, so the diet will be postponed until the day *after* tomorrow. And so on, with the result that the diet never begins.

But what if the present self thought that if the diet didn't begin today, it never would? Or what if the diet had already begun, and the present self knew breaking the diet would set in motion a series of exceptions that would eventually destroy the diet? In that case, the consequences of breaking the diet, or failing to start one, could be sufficiently great that the present self would choose to "be good." And the same goes for tomorrow's self facing the same choice.

This means there's room for a cooperative deal to be struck by all the selves (both the present self and all the future selves that will eventually arrive in the present). They all agree to limit their indulgences. Each self gives up some transient pleasures in return for the restraint exercised by all the others, and on net they are all better off. The agreement is enforced by each self's desire not to destroy the agreement.

The key to Ainslie's intrapersonal bargaining solution is that the cooperative agreement effectively confronts each present self with a "package deal." If the present self weighed the benefit of overeating *just this once* against the small future cost of overeating *just this once*, it would choose to indulge. But the agreement makes it impossible to overeat just once; overeating triggers more overeating. The relevant cost is therefore the cost of overeating *repeatedly*, which is large enough to make the present self abstain.

The agreement just described may sound odd, because we are not accustomed to thinking of ourselves as multiple selves. But in fact, the process is quite common. In enforcing our resolutions, we are loath to make exceptions for fear that they will set a precedent for our own future actions. Successful dieters often adopt rigid personal rules to govern their eating. People trying to quit smoking often do so "cold turkey," because they fear smoking one cigarette will lead to smoking another, and another, until the resolution to quit has been defeated. Wage earners will save a certain amount of money each month, and they will strongly resist reducing the amount—even for just one month—lest they get in the habit of spending more every month.

Personal rules help define the amount of restraint expected of the selves. Bright-line rules, in particular, are valuable as precedents because they can clearly indicate when a present self has chosen to defect from the agreement. Ainslie sees "rationalizations, blind spots, and circumscribed lapse districts" as exceptions to the rules that can "defeat your resolutions."[24] He attributes the problem of backsliding to the temptation to make exceptions.[25] The downside of the collusive solution is that it often results in excessive rigidity. Nonetheless, many people

The self-control problem arises because the present self would like to make a special exception for itself.

The downside of the collusive solution is that it often results in excessive rigidity.

voluntarily adhere to rigid personal rules, presumably because they think the benefits of self-control outweigh the costs.

Establishing Property Boundaries

A key element of the intrapersonal bargaining model just described is that only present selves make decisions. Future selves' interests matter only because the present self happens to care about them (although maybe not as much as it should). But that's not the only way to look at the problem. As an alternative, suppose that the present self does not have exclusive decisionmaking rights. Instead of representing a temporal locus of *control*, the present self represents certain *interests* with a more immediate payoff, while the future self represents interests with a more distant payoff. The two selves exercise joint decision rights over the person. This approach treats the body as a kind of common asset, over which the selves seek to exert control.

In this situation, war is one possible outcome.[26] Each self seeks to advance its own interests while sabotaging the other. The present self searches for chances to overindulge in food, drink, sex, spending, and so forth. The future self finds ways to limit the present self's pleasure—by ridding the household of snacks, throwing cigarettes away, or signing up for automatic savings plan contributions. The future self may also spoil the present self's pleasure by creating guilt or by imposing "oversight" and planning on activities the fun of which derives from their spontaneity.

War is costly to both parties. The present self consumes with attenuated pleasure. The future self's expenditures on enforcement diminish the gains from satisfying its more distant interests. As a result, each self prefers a negotiated outcome. The bargain takes the form of a redistribution of property rights: instead of both selves exercising control at the same time, each self cedes some control over certain kinds of decisions in exchange for exclusive control over others.

Commonplace experience affirms that different interests tend to operate in different circumstances. Individuals adopt rules of self-control, such as, "I will smoke only in social situations," "I will not drink alone," "I will not eat after midnight," "I can ignore my diet while on vacation." Obviously, such personal rules proscribe behavior, and so they are typically interpreted as tools of one's long-run interests. Yet the rules are as notable for what they *allow* as for what they prohibit. Within specified zones, they enable the individual to "let loose" and enjoy life's pleasures without guilt and oversight.

An even better example of an intrapersonal bargain that both constrains and enables is the establishment of separate budgets or accounts for particular activities, such as when a gambler creates a personal gambling fund. Although the fund limits total losses from gambling, it also enables the gambler to gamble freely without worrying about the effects of (sufficiently small) losses on other kinds of consumption. This represents a mutually beneficial exchange between the present and future selves.

Research confirms that people use mental accounts as a means of establishing boundaries.[27] Heath and Soll show that people divide their total resources into "separate mental accounts (e.g., entertainment or household expenses) and then track expenses against the budgets."[28] Wertenbroch observes that people ration their consumption of both "virtue" and "vice."[29] Kivetz and Simonson provide what is likely the best evidence that separate mental accounts *enable* as well as limit consumption: people will deliberately precommit to indulgence by (for instance) choosing luxuries over necessities or cash as lottery prizes.[30]

Here, as in the collusion model discussed above, personal rules assist in the enforcement of an intrapersonal bargain. Unlike that model, the "exceptions" in this model are an integral part of the bargain itself; they are the present self's compensation. The collusion model demonstrates that a certain amount of intrapersonal altruism—the fact that your selves care about each other—can assist in creating rules of self-control. The urge to make exceptions to the rules constitutes a threat. The property rights model, however, shows that some cooperation can occur even if the selves

don't care about each other. In this case, the exceptions don't necessarily threaten the agreement; on the contrary, the exceptions allow the agreement to happen in the first place.

Mutually Beneficial Exchange

If the present and future selves value goods or activities on more than one dimension, then we can imagine yet another kind of intrapersonal bargain. For this kind of bargaining to work, one's selves must possess some form of nonjoint control over the person—either because they have established it through a prior agreement as outlined above or because they possess such control inherently. Suppose, for instance, there are two dimensions of choice: money (present versus future consumption) and food (present indulgence versus future health). And suppose initially the future self has greater control over financial decisions, whereas the present self has greater control over eating decisions. The future self could offer the present self a deal: don't eat that fried chicken, and buy a CD instead. The present self exchanges eating pleasure for listening pleasure. The future self exchanges money (the price of the CD plus interest) for health.

Again, ample evidence supports the idea of intrapersonal exchange. Kivetz and Simonson find that people are most likely to choose luxury rewards for frequent-use programs when they have exerted more effort to obtain the rewards,[31] and they are also more likely to choose luxury rewards when the necessary efforts were related to work rather than pleasure—such as using frequent flier miles for pleasure travel if they were earned via business travel. People also engage in self-gifting to reward themselves for virtuous behavior.[32] Such gifts often perform an "exchange" function by acting as "self-contracts in which the reciprocity for the gift is also personal effort and achievement."[33] Studies have demonstrated the efficacy of self-imposed reward schemes in motivating greater effort and performance.[34]

The three modes of bargaining I have described—collusion, establishment of property rights, and exchange—need not be mutually exclusive, of course. Just as both altruism and self-interest operate between persons, they also both operate within persons. A *limited* degree of intrapersonal altruism could allow for collusive agreements among the selves, while still allowing room for "détente" agreements to avoid costly wars between competing present and future interests. Further opportunities for gain can be exploited via exchanges between the selves.

What Could Possibly Go Wrong?

> In order to carry out a market transaction, it is necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to a bargain, to draw up the contract, to undertake the inspection needed to make sure that the terms of the contract are being observed, and so on.[35]

Given the multiple possibilities for bargaining among one's selves, what might obstruct an efficient and workable outcome? Coase's answer to that question was *high transaction costs*. Transaction costs are the costs (monetary and otherwise) of coordinating the parties to a bargain, negotiating terms, and enforcing the agreement that results.

Transaction costs arise from various sources, but in this context, the most problematic is contract enforcement. Agreements between individuals can be made legally binding by means of explicit contracts enforced by the state legal system. But most intrapersonal agreements must be enforced internally, as the legal apparatus is not usually available. This does not rule out intrapersonal bargaining entirely, but it does mean bargaining selves must depend on mechanisms that are typically less reliable: repeated dealings and reputation. A virtue of the collusion model is that it explicitly incorporates the problem of enforcement, with the solution depending on

There are at least three strong reasons to be skeptical of government interventions designed to fix internality problems.

each self's interest in sustaining cooperation in the future. The other forms of bargaining discussed earlier might be enforced by similar means. If one self persistently violates the terms of its agreements, it signals to other selves its lack of reliability, thus reducing their willingness to make future deals with the violator. The potential violator, realizing this, has reason not to act opportunistically.

The viability of repeated dealings and reputation as modes of contract enforcement depends on the open-ended character of the situation. When the cessation of interaction becomes imminent, "end-game" behaviors can lead to the breakdown of bargaining solutions in both interpersonal and intrapersonal contexts. We might, therefore, expect less self-constraint on the part of people whose lives are coming to an end (though a rational unified self would also engage in greater indulgence under the same circumstances). Also, agreements require adequate policing. John Ameriks, Andrew Caplin, and John Leahy identify "monitoring abilities" as one of the skills that enable households to rein in excessive spending to save more money.[36] Presumably, someone with better monitoring skills can monitor internal agreements at a lower cost.

Although legal enforcement is usually unavailable, other forms of external enforcement do exist. Ainslie refers to such means as "extrapsychic commitments,"[37] a category that includes joining Alcoholics Anonymous and Weight Watchers to enlist the support of other people or advertising one's resolutions to friends and family. Precommitments can also help to enforce contracts by making deviation impossible or very costly. Such commitments include deadlines[38] and automatic savings plan deductions,[39] as well as the tactics mentioned earlier, such as banning fatty foods from the household.

Transaction costs can also arise from the parties' lack of information about each other. A bargainer may, for instance, hold out for a larger share of the gains from trade simply because he thinks the other party values the transaction more than he actually does. In the intrapersonal context, such a problem is less likely. Although it is conceivable that selves may lack perfect knowledge of each other,[40] such knowledge will still be markedly greater than that possessed by different people in an interpersonal context. Hiding or falsification of information cannot be accomplished as easily, given that both selves have access to the same mind.

The policy that effectively corrects some people's problems will fail to correct, or may even exacerbate, others' problems.

Bargaining Failure versus Government Failure

> There is, of course, a further alternative, which is to do nothing about the problem at all. And given that the costs involved in solving the problem by regulations issued by the governmental administrative machine will often be heavy . . . it will no doubt be commonly the case that the gain which would come from regulating the actions which give rise to the harmful effects will be less than the costs involved in governmental regulation.[41]

Given the difficulty of internal contract enforcement, it stands to reason that people do not always succeed in exercising self-control. Some individuals fail at finding effective intrapersonal bargains; they tend to overindulge. Others find imperfect solutions that result in some self-control but not enough. And yet other individuals find solutions that are too effective, resulting in excessive self-control and *under*indulgence. Does it follow that some form of paternalist intervention would correct those problems?

The Coasean perspective argues otherwise. There are at least three strong reasons to be skeptical of government interventions designed to fix internality problems. First, even though some individuals fail to exercise self-control, others succeed. That means internalities are, to some degree, already addressed through intrapersonal bargains. Government interventions could thwart or supersede such bargains. In addition, new bargains would be struck in a different regulatory environment, so we have to ask whether the new bargains would be preferable to the old ones.

Second, interventions have problems of their own. Just as it is incomplete to argue that a market failure alone justifies economic regulation, it is incomplete to argue that a failure of individual choice justifies paternalist regulation. In both cases, the possibility of government failure must be taken into account. Governments often lack the information, the incentives, or both to make wise regulatory decisions.

Third, regulations usually have a "one-size-fits-all" quality, inasmuch as they affect all citizens (though to differing degrees). But people are heterogeneous, meaning the policy that effectively corrects some people's problems will fail to correct, or may even exacerbate, others' problems.

To make these arguments more concrete, we need to consider specific proposals. Here, I will focus on the most obvious and commonly suggested proposal for controlling internalities: the fat tax. More generally, the analysis here will apply to any "sin" tax designed to induce individuals to make better personal health decisions.

Coase + Pigou = Trouble: The Interaction of Taxes and Private Bargaining

Suppose, for simplicity, that the present self makes all decisions about eating, and that the present self cares only about itself. The present self's choice to eat Twinkies creates benefits for itself and imposes costs on future selves. Given these assumptions, a naïve policy analyst would predict that the present self would keep eating Twinkies as long as doing so created any benefits—even if the future costs were very large and the benefits very small. So it might seem like a good idea to tax the present self's eating choices. The tax should be set equal to the future costs of eating Twinkies, so that the present self will take exactly those costs into account.

But this analysis is incomplete, because it ignores the possibility of intrapersonal bargaining. Suppose, for instance, that there are no transaction costs, meaning the selves can reach and enforce internal bargains with little difficulty. In this case, an optimal outcome would occur without any tax. Any time the present self's choice to eat something "bad" would incur greater costs (to the future self) than benefits (to the present self), there is room for a trade. The future self can offer some compensation to the present self, perhaps by offering a reward for abstaining. Since the cost of eating poorly is greater than the benefit, any reward between the two would work: the future self would willingly offer the reward, and the present self would willingly accept.

In this situation, no tax is necessary. But imagine that a tax is imposed anyway, in the mistaken belief that internal bargaining does not occur. Assume that the tax is paid entirely out of the present self's budget. This tax will actually result in *too little consumption*. Suppose the present self could eat a Twinkie, and the present benefit is great enough to justify the future cost. The present self *should* eat. But now the tax diminishes the perceived benefit of eating. The future self will be able to offer a reward to the present self that will induce it to eat this time, whereas without the tax no reward offered by the future self would be large enough. For example, if we measure value in dollar terms, one more Twinkie might be worth $5 to the present self, while causing $4 worth of future damage. This is a Twinkie worth eating, and without the tax it will be eaten. (The future self would not willingly offer more than $4 worth of compensation, and thus the present self would not agree.) But suppose a tax of $1.50 is levied on each Twinkie. Now the present self expects a net benefit of only $3.50, so it will accept a reward offered by the future self not to eat.

This analysis assumes that the tax is paid from the present self's budget. But that may not be true. If the present self does not care (enough) about the future self, why not simply go into debt to pay sin taxes? Incurring debt passes the tax on to the future self. In this case, the future self would perceive an even larger ill effect from the present self's consumption—first the reduction in health and second the

Any one-size-fits-all policy will necessarily be efficient for only a fraction of the public at best.

If internality theory is to be taken seriously, it should incorporate at least some of the lessons learned in the last half century from research on externalities.

reduction in budget. Given the greater cost, the future self would willingly offer larger rewards to induce the present self to reduce its consumption. And once again, the result will be too little consumption. Take the same figures as above: a Twinkie that produces $5 present benefit and $4 future cost. If the future self also expects to experience a $1.50 increase in debt, then it will offer as much as $5.50 for the present self's cooperation—and the present self will accept the offer. The Twinkie does not get consumed, even though it should. In addition, the future self ends up making larger payments to the present self, thus ironically *reducing* the future self's welfare relative to when there was no tax.

I've assumed so far that transaction costs are zero. At the opposite extreme, suppose transaction costs are prohibitively high, so that no intrapersonal bargains are made. Here, the case for a fat tax is stronger. The tax forces the present self to consider the cost to the future self, when otherwise that cost would not have been considered. If the tax comes entirely from the present self's budget, then (as before) the tax effects a welfare shift from the present to the future self. This is the ideal situation for the fat-tax advocate.

On the other hand, if the present self can offload the tax to the future self by going into debt or depleting savings, then the tax has no impact on the present self's consumption. The tax is experienced by the future self as an increase in the cost of present consumption—but by supposition, high transaction costs prevent the future self from making a viable reward offer to the present self to induce it to reduce its consumption. In addition, the tax revenue is lost to the future self. The future self actually ends up worse off, unless the tax revenues are rebated or spent in a way that benefits only the future self.

In reality, transaction costs are neither zero nor prohibitively high. Some internal bargains will be made, others not. The outcome will exhibit elements of both situations just described. The key insight is that some, though probably not all, of the present self's future costs will already have been internalized through intrapersonal bargains. Any tax that fails to account for this process, or to account for it fully, will be too large and thus result in underconsumption by some people. In addition, if the present self can shift taxes to the future self, the policy will tend to diminish the future self's welfare.

Unraveling Intrapersonal Bargains

In the tax analysis above, I treated the bargains between the present and future self as though they were struck in a precise manner, corresponding to exact quantities of consumption. But bargains often take the form of personal rules that divide up or reallocate decisionmaking power. Instead of specifying the number of fat calories the present self may consume, the rule might specify times and places at which the present self may freely consume fat and other circumstances in which it may not.

The effects of a fat tax on idiosyncratic bargains of this kind are more difficult to parse. In the short run, existing personal rules will likely persist. Especially when transaction costs are high, bargainers have an interest in maintaining existing agreements to economize on such costs and avoid a breakdown in the relationship, even if those agreements are no longer optimal. The present self may continue eating fatty foods only on weekends and vacations, for example. This could occur even though it would make sense, given the tax, to shrink the set of allowed indulgence zones. If the present arrangement is already optimal or nearly so, such persistence could be desirable. If the individual had not succeeded in reaching internal bargains for self-control, then the persistence of old rules would be undesirable, though the tax would not aggravate the situation (except by reducing the selves' income).

Eventually, however, people will renegotiate their internal bargains. To minimize the tax's impact, they will find it worthwhile to reduce their level of consumption. They will try to find a new set of personal rules that

approximate the desired level of consumption, which may be difficult to do (perhaps indulgence is allowed only every other weekend or only on Sundays). In any case, whatever new rules appear might or might not improve overall welfare. To the extent that the tax falls on only the present self's income *and* transaction costs prevent the negotiation of efficient personal rules in the absence of the tax, the tax will tend to induce better personal rules. But if transaction costs are low enough that the selves eventually tend to arrive at near-efficient rules absent the tax, the tax will tend to reduce consumption below the optimal level. And if transaction costs are high while the present self can offload the tax to the future, the tax will reduce the future self's income while failing to reduce the present self's consumption.

Information and Incentives

Given the difficulties outlined above, would-be paternalist regulators face a daunting task. Even without Coasean considerations, optimal taxation of internality-producing behavior would be no simple task. The optimal tax would be equal to the marginal cost of the behavior to future selves. To calculate this amount, regulators would first need to find the "true" rate of tradeoff between present and future satisfaction. The problem, as noted earlier, is that the "true" rate of time-discounting is a phantom. To pick one rate of tradeoff over another is to privilege one set of subjective preferences over another, without any basis for doing so.

But suppose the regulators somehow found the "right" rate of tradeoff between the present and the future. Even then, they would need to discern the degree to which people have already dealt with their internalities through intrapersonal bargaining. Such information will not be readily available. The phenomenon of time inconsistency has been identified primarily under laboratory conditions, in which test subjects are presented with stylized choice situations (e.g., "Would you prefer $100 now or $110 tomorrow? Would you prefer $100 a year from now or $110 in a year and a day?"). The rates of tradeoff over time revealed by such experiments will not necessarily, or even likely, approximate the rates of tradeoff used by people in real-world situations. The actual devices people use to define and enforce intrapersonal bargains, and thus to induce more future-oriented behavior, most often involve personal rules based on circumstances (e.g., "Am I in a bar right now? Am I on vacation?") that do not appear in the laboratory setting.

Moreover, even if regulators could discern both the "right" rate of time tradeoffs and the actual rates of tradeoff implicit in people's behavior, they would still face the unenviable task of estimating the degree to which subsequent choice by the regulated people will undermine their policies' intended results. Since people may change their choice process in response to policy changes—for example, by altering the terms of their internal agreements—it follows that realized rates of tradeoff will be endogenous to the policy choice.

Furthermore, people are heterogeneous—in the size of their initial internality problems, the magnitude of their internal transaction costs, and the type of personal rules available and attractive to them. Any one-size-fits-all policy will necessarily be efficient for only a fraction of the public at best. Others will be unaffected or affected adversely by being manipulated into suboptimal consumption, or affected too little because the policy doesn't go far enough. Any attempt to improve the policy's effectiveness vis-à-vis the latter group will have undesired and often unexpected consequences for the other groups.

And with all of the informational difficulties, we have not even begun to ask whether regulators will have the appropriate incentives to find the correct answers and implement them.

Conclusion

Does the theory of internalities justify government intervening in people's lives "for

There's no valid reason to assume, when there is an inconsistency between present and future interests, that the latter must trump the former.

Individuals have every reason to understand their own needs and find suitable means of solving their own problems.

their own good"? The new paternalists clearly think so. But their argument is extraordinarily weak.

The theory of internalities is explicitly modeled on the theory of externalities. If internality theory is to be taken seriously, it should incorporate at least some of the lessons learned in the last half century from research on externalities. But that hasn't happened yet. As it stands, the case for paternalism based on internality theory suffers from several major flaws.

First, the new paternalism blithely assumes that, when your present self can impose costs on your future self, the outcome is necessarily bad. But preventing harm to the future self might involve even greater harm to the present self. There's no valid reason to assume, when there is an inconsistency between present and future interests, that the latter must trump the former.

Second, the new paternalism ignores the fact that harms can be avoided in multiple ways. Restricting present behavior is one way to reduce future harms, but that doesn't make it the *best* way. The future self might be capable of mitigating the harm at lower cost by other means.

Third, the new paternalism neglects the possibility of internal bargains and private solutions. All of us face self-control problems from time to time. But we also find ways to solve, or at least mitigate, those problems. We make deals with ourselves. We reward ourselves for good behavior and punish ourselves for bad. We make promises and resolutions, and we advertise them to our friends and families. We make commitments to change our own behavior. Internality theorists point to these behaviors as evidence that the internality problem *exists*. But they are actually evidence of the internality problem being *solved*, at least to some degree.

People are not perfect, so we should not expect real people's actions to mimic those of perfectly rational and perfectly consistent beings. Mistakes will occur; self-control problems will persist. But paternalist solutions will solve them no better than personal solutions. What is really at stake is *how* self-control problems will be addressed—through private, voluntary means or through the force of government.

The new paternalists would have us believe that benevolent government can—through taxes, subsidies, restrictions on the availability of products, and so on—make us happier according to our own preferences. But even if we place little or no value on freedom of choice for its own sake, the paternalists' recommendations simply don't follow. Public officials lack the information and incentives necessary to craft paternalist policies that will help the people who most need help, while not harming those who don't need the help or who need help of a different kind. Individuals, on the other hand, have every reason to understand their own needs and find suitable means of solving their own problems.

Notes

1. This is a simplification, since some new paternalists also draw on arguments about individuals' lack of information or poor information-processing skills. Here, I will set aside those related but distinct arguments to focus on the problem of intrapersonal conflict and choice.

2. "I say only that people act as if there were two selves alternately in command . . . the ways in which people cope, or try to cope, with loss of command within or over themselves are much like the ways that one exercises command over a second individual." Thomas Schelling, "Ethics, Law, and the Exercise of Self-Command," in *Choice and Consequence: Perspectives of an Errant Economist*, ed. Thomas Schelling (Cambridge, MA: Harvard University Press, 1984), p. 84.

3. Ronald H. Coase, "The Problem of Social Cost," *Journal of Economic Literature* (October 1960).

4. A. C. Pigou, *The Economics of Welfare*, 4th ed. (London: Macmillan, 1932).

5. Jonathan Gruber and Botond Koszegi, "Is Addiction 'Rational'? Theory and Evidence," *Quarterly Journal of Economics* 116 (2001): 1261–94.

6. Ted O'Donoghue and Matthew Rabin, "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes," *American Economic Association Papers & Proceedings* 93 (2003): 186–91.

7. Ted O'Donoghue and Mattew Rabin, "Optimal

Sin Taxes," unpublished manuscript, University of California at Berkeley, 2003, p. 2, n. 3.

8. Colin Camerer et al., "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism,'" *University of Pennsylvania Law Review* 151 (2003): 1211–54.

9. Richard H. Thaler and Cass R. Sunstein, "Libertarian Paternalism," *AEA Papers and Proceedings* 93 (2003): 175–79.

10. Camerer et al., p. 1214.

11. Ibid., p. 1221.

12. Coase, p. 2.

13. See, for instance, Richard H. Thaler and H. M. Shefrin, "An Economic Theory of Self-Control," *Journal of Political Economy* 89 (1981): 392–406. See also Jon Elster, "Weakness of the Will and the Free-Rider Problem," *Economics and Philosophy* 1 (1985): 231–65.

14. Ran Kivetz and Itmar Simonson, "Self-Control for the Righteous: Toward a Theory of Precommitment to Indulgence" *Journal of Consumer Research* 29 (2002): 199–217; and Dana N. Lascu, "Consumer Guilt: Examining the Potential of a New Marketing Construct," in *Advances in Consumer Research* 18 (1991): 290–95. These scholars, among others, document the importance of guilt as a motivator in decisionmaking.

15. See, for example, Tyler Cowen, "Self-Constraint versus Self-Liberation," *Ethics* 101 (1993): 360–73; and George Ainslie, *Breakdown of Will* (New York: Cambridge University Press, 2001), p. 115.

16. See Chip Heath and Jack B. Soll, "Mental Budgeting and Consumer Decisions," *Journal of Consumer Research* 23 (1996): 40–52.

17. Mathematically, time inconsistency follows from a quasihyperbolic utility function such as this one:

$$U^t(u_t + u_{t+1}, \ldots, u_T) = u_t + \beta \sum_{s=t+1}^{T} \delta^s u_s$$

The discount factor β, which is less than one, represents the agent's degree of present bias. Since β does not apply to the present period (time t), the tradeoff between any two periods will depend on whether one of those periods is the present. As a result, the tradeoff changes over time instead of being consistent. If β equals one, the agent is perfectly consistent.

18. The two "corrections" can be seen using the quasihyperbolic utility function above. One correction involves setting β equal to one; this favors future selves. The other correction involves applying β to every period instead of just the present (in effect, lowering δ to $\delta\beta$); this favors the present self. With either correction, the person's behavior will become consistent over time.

19. Coase, p. 41.

20. See, for example, *Boomer v. Atlantic Cement Co.*, 26 N.Y.2d 219, 257 N.E.2d 870; 309 N.Y.S.2d 312 (NY Court of Appeals 1970).

21. I ignore, for the sake of simplicity, the possibility of the factory paying the residents to move away in expectation of a reduced tax burden.

22. Coase, p. 15.

23. George Ainslie, *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person* (Cambridge: Cambridge University Press, 1992).

24. Ibid., p. 189.

25. George Ainslie, "How Do People Choose between Local and Global Bookkeeping?" *Behavioral and Brain Sciences* 19 (1996): 574–75.

26. Consider Thomas Schelling's notion of limited war, which George Ainslie describes as the relationship of "bargaining agents who have some incompatible goals but also some goals in common." George Ainslie, "A Research-Based Theory of Addictive Motivation," *Law and Philosophy* 19 (2000): 100.

27. See Richard H. Thaler, "Mental Accounting and Consumer Choice," *Marketing Science* 4 (1985): 199–214.

28. Heath and Soll, p. 40.

29. Klaus Wertenbroch, "Consumption Self-Control by Rationing Purchase Quantities of Virtues and Vice," *Marketing Science* 77 (1998): 317–37.

30. Kivetz and Simonson, "Self-Control for the Righteous."

31. Ran Kivetz and Itamar Simonson, "Earning the Right to Indulge: Effort as a Determinant of Customer Preferences toward Frequency Program Rewards," *Journal of Marketing Research* 39 (2002): 155–70.

32. David Glen Mick and Michelle DeMoss, "Self-Gifts: Phenomenological Insights from Four Contexts," *Journal of Consumer Research* 17 (1990): 322–32; David Glen Mick, "Giving Gifts to Our-

selves: A Greimassian Analysis Leading to Testable Propositions," in *Marketing and Semiotics: Selected Papers from the Copenhagen Symposium*, ed. Hanne Hartvig Larsen, David Glen Mick, and Christian Alsted (Copenhagen: Hendelshojslolens, 1991), pp. 142–59; and David Glen Mick, "Self-Gifts," in *Gift Giving: A Research Anthology*, ed. Cele Otnes and Richard F. Beltramini (Bowling Green, OH: Bowling Green State University 1996), pp. 99–120.

33. Mick and Demoss, p. 326.

34. See, for example, Albert Bandura and Bernard Perloff, "Relative Efficacy of Self-Monitored and Externally Imposed Reinforcement Systems," *Journal of Personality and Social Psychology* 7 (1967): 111–16; and Albert Bandura and Dale H Schunk, "Cultivating Competence, Self-Efficacy, and Intrinsic Interest through Proximal Self-Motivation," *Journal of Personality and Social Psychology* 41 (1981): 586–98.

35. Coase, p. 15.

36. John Ameriks, Andrew Caplin, and John J. Leahy, "Wealth Accumulation and the Propensity to Plan," *Quarterly Journal of Economics* 118 (2003): 1007–47.

37. George Ainslie, *Breakdown of Will*, pp. 74–76.

38. Dan Ariely and Klaus Wertenbroch, "Procrastination, Deadlines and Performance: Self-Control by Precommitment," *Psychological Science* 13 (2002): 219–24.

39. Richard H. Thaler and Shlomo Benartzi, "Save More Tomorrow ™: Using Behavioral Economics to Increase Employee Savings," *Journal of Political Economy* 112 (2004): S164–S182.

40. See, e.g., Roland Bénabou and Jean Tirole, "Will Power and Personal Rules," *Journal of Political Economy* 112 (2004): 848–86, for a model of self-control that incorporates imperfect recall.

41. Coase, p. 18.

CATO INSTITUTE
1000 Massachusetts Ave., N.W.
Washington, D.C. 20001

Published by the Cato Institute, Policy Analysis is a regular series evaluating government policies and offering proposals for reform. Nothing in Policy Analysis should be construed as necessarily reflecting the views of the Cato Institute or as an attempt to aid or hinder the passage of any bill before Congress. Contact the Cato Institute for reprint permission.

Additional copies of Policy Analysis are $6.00 each ($3.00 each for five or more). To order, or for a complete listing of available studies, write the Cato Institute, 1000 Massachusetts Ave., N.W., Washington, D.C. 20001 or call toll free 1-800-767-1241 (noon–9 p.m. eastern time).
Fax (202) 842-3490 • www.cato.org

CATO INSTITUTE