2009

Willpower and Legal Policy

Lee Anne Fennell

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation


This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.
WILLPOWER AND LEGAL POLICY

Lee Anne Fennell

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

June 2009

Willpower and Legal Policy

Lee Anne Fennell

Introduction

People often act in ways that are inconsistent with their own stated desires, whether by overeating, undersaving, procrastinating, overworking, compulsively gambling, shopping, drinking, or smoking, or otherwise failing to live up to some self-imposed ideal in activities or consumption patterns. What, if anything, can or should legal policy do about this disjunction? In recent years, the law and social science literature has increasingly recognized the potential relevance of willpower and related concepts—self-control, impulsiveness, time-inconsistent preferences, myopia, hyperopia, and so on—to social policy. A growing body of scholarship examines the workings and implications of willpower in areas such as personal and public finance (e.g., Laibson 1996, 1997; Laibson et al. 1998; Diamond & Köszegi 2003; Kumru & Thanopulos 2008), consumer contract design (e.g., DellaVigna & Malmendier 2004; Bar-Gill 2004), health and personal well-being (e.g., Gruber & Köszegi 2001; Herman & Polivy 2003; Strnad 2005; O’Donoghue & Rabin 2006; Gul & Pesendorfer 2007; Dodd 2008), and criminal law (e.g., Wilson &

To take a simple example, suppose Alice faces an onerous task like revising a manuscript. She professes at Time 1 that she strongly prefers to do the task at Time 2 and then take it easy at Time 3. When Time 2 arrives, however, we might find Alice relaxing rather than working, perhaps stating a new desire to work at Time 3. Alternatively, after sticking to a plan of manuscript revision at Time 2, we might find Alice undertaking some additional burdensome task at Time 3, rather than relaxing as she had planned. In either or both of these scenarios, Alice expresses time-inconsistent preferences, and questions of willpower are potentially implicated. The potential policy relevance of such seemingly trivial intrapersonal dramas becomes apparent if we suppose the population is filled with people like Alice with willpower issues that make them chronically worse off along dimensions such as health, family interactions, finances, education, and employment.

Pinning down the implications of willpower for legal policy requires clarity on three fronts. The first is definitional: because self-identified gaps between ideal and actual behavior represent only a subset of the ways in which choice over time can go awry, willpower problems must be carefully isolated from other, related phenomena. Second, because willpower or self-control problems imply an intraself conflict, the terms of that conflict require clear specification. A particular challenge in this regard is in deciding which “self” ought to be privileged through policy choices (see, e.g., Feinberg 1986, 83; Whitman 2006, 5; Bernheim & Rangel 2008, 156), an inquiry that in turn requires some criterion for assessing well-being (see, e.g., Bernheim & Rangel 2008; Loewenstein & Ubel 2008). Third, the various reasons policymakers might care about willpower problems
must be carefully disaggregated from each other and from concerns that are not, at bottom, related to self-control issues. This requires identifying and examining the costs associated with willpower lapses and their prevention. The first three parts of this review are organized around these three areas of analytic concern, all of which have received significant recent attention from scholars. In the last two parts of the review, I briefly survey the literature on policy alternatives and identify areas for future research. This short review cannot begin to provide a comprehensive account of the vast and rapidly proliferating scholarship on this topic, but I hope that it will convey some of the richness, variety, and energy surrounding this area of research.

I. Defining and Isolating Willpower

Although definitions vary, “willpower” is perhaps best understood as a synonym for “self-control”—an individual’s efforts to resist a presently attractive course of conduct in favor of a self-identified alternative that she deems superior. The fact that self-control failures are intensely subjective phenomena – behavior falls short of a benchmark that the individual has chosen for herself – complicates the relationship among willpower, observable behavior patterns, and social value. Because willpower can be deployed in pursuit of any goal an individual chooses, willpower failures need not produce socially disfavored conduct. For example, self-control deficits can generate “hyperopic” behaviors of which many would approve, such as workaholism or oversaving (see, e.g., Elster 1985, 6, citing Davidson 1980, 30; Kivetz & Simonson 2002; Hamermesh & Slemrod 2008). To

1 Usages vary. See, e.g., Baumeister & Vohs 2003, 202-04 (using “willpower” to refer to a particular theory of self-control).
take a more extreme example, society is better off if villains suffer from a weakness of will in carrying out antisocial acts (see Utset 2007, 665-68; Utset 2009).

By the same token, intertemporal decisions that have unwanted personal or societal consequences are not always the product of willpower shortfalls. Behavior that is observationally equivalent to that produced by low willpower can stem from unconflicted time preferences or from mistakes and miscalculations, as the following sections explain. Recognizing this point is crucial because choices generated by different kinds of cognitive phenomena can point in divergent, even opposite, policy directions (see, e.g., Köszegi & Rabin 2008, 206-07, 208 n.7).

A. Conflicted and Unconflicted Time Preferences

Because all choices are situated in time, the notion of conflicted time preferences has proved essential in identifying the subset of intertemporal decisionmaking that implicates willpower. Significantly, choices over time are often driven by considerations wholly unrelated to time preferences, such as liquidity constraints (e.g., Deaton 1992, 162-63), uncertainty about the future in a world with incomplete insurance markets (e.g., id. at 34-37; Kelman, 1983, 660-70), changing circumstances over the life cycle (e.g., Deaton 1992, 5), or preferences about consuming in particular lumps or patterns (e.g., Shaviro 2007, 766; Loewenstein & Prelec 1993). Similar complications confound experimental efforts to examine time preferences. As Frederick et al. (2003, 61) explain, “spectacular disagreement among dozens of studies that all purport to be measuring time preference” may be explained by “the fact that the various elicitation procedures used to measure time preference consistently fail to isolate time preference,” and instead mix in elements such as
“intertemporal arbitrage, when tradeable rewards are used; concave utility; uncertainty that the future reward or penalty will actually obtain; inflation, when nominal monetary amounts are used; expectations of changing utility; and considerations of habit formation, anticipatory utility, and visceral influences” (ibid.; see also ibid., 50-62).² Even when time preferences are squarely implicated, self-control may not be. A person with a consistently high discount rate, for example, does not experience the internal conflict that is the hallmark of willpower. A rationally maximizing spendthrift, overeater, or drug addict (see Becker & Murphy 1988) may have a problem, but it is not a self-control problem.

Inconsistent time preferences, in contrast, are typically associated with self-control problems—although they need not be as a matter of logical necessity. Formal study of “dynamic inconsistency” in time preferences began over half a century ago (Strotz 1955), although earlier discussions of self-control problems can certainly be found (see, e.g., Palacios-Huerta 2004, 75-78 (discussing Smith 1759)). It is now well established that people do not always discount future payoffs at a constant rate (see, e.g., Loewenstein 1992; Frederick et al. 2003). For example, experiments suggest that both animals and people can experience preference reversals as a pair of temporally separated choices draws nearer in time, even though the relative values of the payoffs and the temporal distance between them has not changed (see e.g., Ainslie 1992, 78). Hence, Prudence might prefer $105 after 366 days to $100 after 365 days, but change her mind when the choice becomes one between $100 today and $105 tomorrow (see, e.g., Frederick et al. 2003, 25). Such

² Of course, it is often controversial what should and should not be considered integral to time preference. Some difficulties stem from inconsistent use of terms, or the sheer profusion of terms used to describe a variety of similar, overlapping, and distinct time-related concepts (see Frederick et al. 2003, 61-62 & fig. 1.4; ibid., 73 n. 42 (listing 19 different terms used in discussing choice over time)).
preference reversals have generally been explained as the product of hyperbolic
discounting (see ibid.), although some scholars have offered competing explanations (see,
e.g., Read 2001; Rubenstein 2003).

For purposes of identifying willpower problems, the precise reason for the
preference reversal is less important than the internal experience preceding it. Intuition and
self-reflection suggest that an internal conflict will typically precede the reversal, as the
long-run self loses ground relative to the demands of a present self. But it is not
unimaginable, at least at the level of theory, that preference reversals could occur over time
in a manner that is wholly free of internal conflict. To continue with the example above,
perhaps Prudence has a metapreference (see Becker 1996, 132) for always choosing in
accordance with her present desires. Preferences expressed at a point removed from the
moment of choice, when there is no then-present interest on the line, are consistently
subordinated within her decision structure to the later preferences of then-current selves.
Thus, her preference reversals do not stem from the internal conflict associated with the

---

3 With hyperbolic discounting, there is a dramatic drop-off in the value an individual assigns to a
reward as it is delayed from the present to the near future, but as the reward becomes increasingly
distant, its present value declines at a decreasing rate; the curve flattens out (see, e.g., Ainslie 2001,
32 & fig. 2B). Thus, a hyperbolic discounter comparing two rewards of different sizes that are
offered one day apart may find the larger, later reward worth waiting for when that one-day gap is
situated in the relatively distant future, in the flat part of the curve for both rewards’ hyperbolas
( ibid.), but change her mind when comparing the earlier reward’s immediate availability (at its full,
undiscounted value) with a later reward whose value falls in the precipitously declining portion of
its hyperbola (see ibid). Such preference reversals have been formally modeled using a quasi-
hyperbolic discount function originally formulated by Phelps & Pollak (1968) for use in the
intergenerational setting. See, e.g., Elster 1979, 71-77; Laibson 1997; Frederick et al. 2003, 32-34
(reviewing literature). See also Angeletos et al. 2003, 536 (finding based on simulations that “[a]ll
in all, a model of consumption based on a hyperbolic discount function consistently better
approximates the data than does a model based on an exponential discount function”).

4 As discussed below, a two-self model is often used to capture the internal experience associated
with willpower exertions (see, e.g., Thaler & Shefrin 1981).
unsuccessful application of willpower, but rather from application of a stable metapreference to newly available information about the present self’s desires.

B. The Problem of Faulty Targets

Willpower lapses must also be conceptually disaggregated from cognitive errors that keep people from recognizing what is best for them to do, although the two kinds of problems are entangled with each other (see Bernheim 1999, at 36, 38). Computational limits, incorrect or incomplete information, and uncertainty about the future may cause people to aim their willpower efforts at a faulty target. For example, a “projection bias” often limits people’s capacities to predict how they will feel if circumstances change in particular ways (Loewenstein et al. 2003a; see Wilson & Gilbert 2003). When people change their minds, then, the switch might either be a new mistake (an impulsive self derailing optimal plans) or evidence of an old mistake (a prior self not understanding what would, in fact, be optimal in a given setting). Köszegi and Rabin (2008, 206-07) illustrate this point with the example of a woman who wishes to deliver a child without using anesthetics, but then requests anesthetics while in labor. Without knowing whether this request stems from a failure of willpower or, instead, from newly available information about actual pain levels, we cannot know whether honoring it would maximize her welfare (ibid.).

Miscalculations can also combine or interact with willpower problems to reduce well-being. For example, if an individual concocts a long-range savings plan that falls short of optimality and then fails even to live up to that inadequate goal, only the latter failing can be properly chalked up to willpower problems. Conversely, if an individual
fails to live up to an overly ambitious savings plan that prescribes less spending than is optimal, her failure to meet her self-imposed goals may be only partly attributable to low willpower. As will be developed further below, self-control is costly to exert (e.g., Baumeister & Vohs 2003) and lapses often beget further lapses (Ainslie 1985, 147; Bénabou & Tirole 2004, 851-56; Bénabou & Tirole 2003, 153-56). Had the individual chosen a more realistic savings goal that was closer to the optimum, she might well have been able to exercise the necessary level of self-control. Here, the willpower failure has become entwined with other difficulties. Imperfections in memory present a similar complication. If it is too difficult to keep track of the optimal consumption pattern, a suboptimal but more memorable pattern may be chosen instead (Gilboa & Gilboa-Schechtman 2003, 132-34).

People may also misgauge their own future susceptibility to self-control problems (O’Donoghue & Rabin 1999, 2003a). For example, research supports the existence of a “hot/cold empathy gap” (Loewenstein 2000a, at 428-29) that causes people to underestimate the extent to which visceral factors will influence their decisionmaking. While a lack of insight into one’s own future propensities is not itself a failure of willpower, it can complicate efforts to address willpower problems. If people are partially or fully naive about their willpower shortcomings, they will be less inclined to pursue strategies, such as precommitment, that will compensate for those limitations (see e.g., O’Donoghue & Rabin 1999, 2003a).

II. The Battle of the Selves
Self-control implies an internal struggle over intertemporal choice; it assumes at least the intermittent presence of “two selves” vying for decisionmaking power (e.g., Loewenstein 2000b, at 52; Elster 1985, at 6; Thaler & Shefrin 1981, 393-94, quoting McIntosh 1969). For example, Thaler & Shefrin (1981) frame self-control problems as pitting the long-run preferences of a “planner” self against the immediate, short-run desires of a “doer” self. This approach raises questions not only about the two-self formulation, but also about how the disagreement between the selves should be conceptualized.

A. The Two-Self Metaphor

The idea that more than one “self” fights for control over an individual’s decisions is clearly metaphorical (e.g., Loewenstein 1996, 288). An initial question is why scholars studying self-control problems would choose to employ such a construct. The law generally assumes that each individual constitutes a single self, and assigns responsibility on that basis (see, e.g., Feinberg 1986, 83-84; Schelling 1984, 96-97). Even if we recognize that certain aspects of an individual’s identity may change over time, we might nonetheless regard her as acting, at each moment, as an optimizing single owner of all her lifetime resources, or as a faithful trustee for all incarnations of herself. This conception of the individual implicitly animates the standard economic models for consumption over the life cycle (see, e.g., Zelenak 2009; see also Friedman 1957, 25-31 (presenting the “permanent income hypothesis”); Modigliani & Brumberg 1954 (presenting the “life-cycle model”)).
Self-control models challenge the picture of the self as an optimizing single owner, but they do not discard the assumption of a stable, continuing identity.\(^5\) Indeed, the premise of continuity is implicit in the choice to model willpower problems using just two contemporaneous selves rather than a series of temporal selves independently governing each successive segment of an individual’s life. The essential interaction might be understood as playing out between a “present self” (a position occupied by a succession of temporal entities) and a “long-run self” who, at least in theory, can act as a fiduciary for the person’s future interests (see Zelenak 2009, n.150). Consistent with this interpretation, Shefrin & Thaler (1988, 611-12) suggest that the selves have “contrasting time horizons.”\(^6\)

The two selves might also be cast in psychological terms corresponding to Freudian or Platonic distinctions (McIntosh 1969, 122-25). Work in neuroeconomics suggesting that different brain structures are activated by prospects of immediate gratification and by longer-range considerations, respectively (e.g., McClure et al. 2004; Benhabib & Bisin 2005), offers another potential way of characterizing the conflict between selves (see Manuck et al. 2003, 162-63). However, recent research has challenged such “dual systems” approaches (Johnson 2008, 15). For example, Glimcher et al. (2007, 143) found “no evidence of separable neural agents that could account for the multiple selves that are used to explain hyperbolic-like discounting behavior” in their neuroeconomic study of intertemporal choice. While the two-self idea should not be taken too literally, it can nonetheless help to highlight the conflicts with which legal policy might be concerned.

\(^5\) For an extended philosophical exploration of personal identity over time, see Parfit 1984.

\(^6\) On one account, the “long-run self” has interests that are just as biased as those of the “present self,” albeit in a different direction. See Posner 1997, 25. It is also possible to treat the “long-run self” as a trustee charged with optimizing over the entire life cycle, including the present episode. Cf. O’Donoghue & Rabin 1999, 113 (formulating a “long-run perspective” that weights each period equally).
B. Intrapersonal Transaction Cost Environments

A primary payoff to framing self-control problems as conflicts among selves is that doing so facilitates useful analogies to interpersonal interactions. In the interpersonal realm, costs that an actor does not take into account in making a decision—externalities—can result in inefficiency. Herrnstein et al. (1993, 150) adapted the notion of externalities to the intrapersonal context by terming unaccounted-for costs that one self imposes on other temporal selves “internalities.” As Coase (1960) demonstrated, in a zero-transaction cost environment, all costs would be taken into account and an efficient result would always be reached, regardless of the starting points embedded in a given legal regime. In real world environments, whether or not an actor will take a given cost into account depends not only on the prevailing legal rules but also on transaction costs. These same basic points have recently received attention in the literature on intrapersonal dilemmas (Whitman 2006; Rasmusen 2008; Dodd 2008).

The existence of self-control problems suggests an internal bargaining environment that has positive transaction costs (see Whitman 2006, 6-13) and in which failures in contracting are possible (see, e.g., ibid; Brocas et al. 2004, 60-63). The dilemmas resemble in some respects those faced by owners of temporally adjacent interests, such as a life estate and a remainder (see Posner 2007, 74-75). There, the literal unavailability of a

---

7 Externalities only produce inefficient results when they lead actors to make different decisions than the ones they would have reached after taking external costs and benefits into account (Buchanan & Stubblebine 1962); this will not always be the case (ibid.; Haddock 2007). Nonetheless, unaccounted-for costs tend to distort decisionmaking and produce inefficiencies (Dukeminier et al. 2006, 44-45).

8 The bargaining in question may be largely implicit (see Ainslie 2001, 105-16), although certain strategies, such as those involving commitments to reward oneself (see Koch & Nafziger 2009), seem to involve more explicit dealmaking.
party whose interests are affected by present-day choices (such as an unborn descendant who holds a future interest) can make bargaining impossible (ibid., 74). While the very existence of a willpower problem suggests that the future selves are not wholly and chronically unrepresented, bargaining may still be quite costly (e.g., Schelling 1984, 94). The long-run self may suffer from short-run interruptions in her ability to communicate persuasively to the present self. Especially strong present desires, such as those triggered by visceral factors (see, e.g., Loewenstein 1996; Loewenstein 2000b, 52-55) might offer one explanation for these communication breakdowns. Another possibility, explored below, relates to the effects of previous cognitive and self-control exercises on one’s present ability to engage in self-regulation.

C. Which Self Gets Priority?

The analogy to interpersonal interactions also raises another issue: which “self” should get priority in the event of a conflict? (e.g., Schelling 1984; Whitman 2006, 4-5; Andreou 2008, 29-30). The question fits within a larger controversy about the appropriate role of new forms of paternalism, whether termed “soft” (e.g., Feinberg 1986, 12); “light” (e.g., Loewenstein & Haisley 2008), “asymmetric” (Camerer et al. 2003), “cautious” (O’Donoghue & Rabin 1999, 150), or “libertarian” (Sunstein & Thaler 2003). Behavioral law and economics increasingly suggests that we cannot simply rely on revealed preferences as a guide to what will make people better off—even in their own estimation (e.g., Bernheim & Rangel 2008). The possibility that people are actually making themselves worse off in their own eyes has led to new calls for public policy interventions, cast as efforts to help individuals achieve outcomes they prefer (e.g., Camerer et al. 2003;
Thaler & Sunstein 2008). Yet indeterminacy as to what best serves an individual’s own utility function (see Korobkin 2009) has fueled criticism of the new paternalist program (e.g., Mitchell 2005; Glaeser 2006; see also Loewenstein & Haisley 2008, 212-15 (reviewing critiques)).\footnote{Another response to indeterminacy would be to reject the subjective utility of the chooser in favor of more objective measures of well-being (see, e.g., Lewinsohn-Zamir 2003). For example, one could assess the desirability of a social policy by examining its effects on human capabilities, drawing on the work of Amartya Sen and Martha Nussbaum (see Loewenstein & Haisley 2008, 220-21), although such approaches present difficulties of their own (see ibid). Korobkin (2009) proposes a “libertarian welfarist” approach that shifts attention from the chooser’s utility to the externalities associated with the choice. This approach does not solve the indeterminacy problem, but rather suggests that certain measures to influence choice should be taken regardless of the impact on the chooser’s utility when external effects are large and unidirectional (see ibid., 37-41).} Time-inconsistent preferences engage this issue in an especially pointed manner, because an individual’s own stated preferences conflict with each other.

As Bernheim and Rangel (2008, 156) recently observed, much work on self-control and time-inconsistent preferences assumes that the longer-range self’s preferences more closely track welfare optimization. But because a time-inconsistent preference could be explained by an error on the part of the long-run self as easily as it could be explained by an error on the part of the short-run self, favoring the long-run self’s choice appears arbitrary (see, e.g., Besharov 2004, 13; Whitman 2006). A different approach would be to examine the decisionmaking process to see whether a deviation from the assumption that people are making welfare-maximizing choices seems warranted. In this vein, Bernheim and Rangel (2008, 186-89) have suggested that “nonchoice evidence” that information-processing has been impaired in a given circumstance can be used to identify “suspect” choices that should be disregarded. Advances in neuroeconomics may eventually make more transparent the brain processes that produce time-inconsistent preferences so that processing mistakes can be more easily detected (see ibid., 189), but we are presently far
from being able to reliably specify what constitutes a malfunction (Loewenstein & Ubel 2008).

The transaction cost analysis discussed earlier suggests we might approach the problem in yet another way—not by attempting to spot malfunctions, but rather by establishing procedural safeguards that facilitate self-correction of errors. Suppose, for example, that we can identify situations where the longer-run self has little time or ability to confer with the present self before a decision is made. Even if we cannot be certain that the present self is making a mistake, policy might attempt to alter the decisionmaking environment in ways that reduce impediments to intrapersonal bargaining. Perhaps for this reason, waiting or “cooling-off” periods are familiar policy tools, despite the inconvenience they may cause (e.g., Cooter 1991, 158; see Loewenstein & Haisley 2008, 239 n. 7). While such mechanisms raise interpersonal transaction costs, they could lower internal transaction costs for some individuals.

III. The Costs of Self-Control Problems

Because legal policy would have no interest in self-control problems unless some kind of harm stemmed from those problems, it is important to specify the nature of that harm. Significantly, costs can flow not only from willpower lapses themselves but also from efforts directed at keeping them from happening (cf. Calabresi 1970). The literature has identified three categories of negative impacts, which I will here term “failure costs,” “exercise costs,” and “erosion costs.”

---

Although I have not seen this precise taxonomy used elsewhere, the categories I describe are well known in the literature.
A. Failure Costs

Successful applications of willpower allow people to choose differently than they would based on immediate preferences (see, e.g. Loewenstein 2000b). These efforts are presumably undertaken to avoid some negative result that would be associated with the more immediately appealing alternative—that is, the costs associated with a failure of willpower. In assessing such “failure costs,” it is necessary to examine both the effects of the lapse-induced choice on the individual’s well-being and the effects of that choice on other people and society at large.11

To take a simple example, suppose Austin eats a piece of cake that he tried unsuccessfully to resist eating. This slip may move Austin further from the consumption path that is optimal for him (in his own estimation), making him less well off. Of course, it is possible that the lapse actually improves Austin’s lifetime well-being if, for example, the exertion of willpower was prompted by a mistaken projection about the future or misinformation about the applicable health risks.12 The effect of the willpower failure on Austin’s well-being may be amplified or offset by the effects of his decision on other people and on society as a whole. For example, if the cake-eating episode contributes to health declines in Austin, others in his health insurance pool may bear some of the

11 It is possible that “failure costs” could be nonexistent or even negative, as where people apply willpower toward ends that actually subtract value from themselves or others. See, e.g., Utset 2007, 668–71.
12 The fact that choices are based on inaccurate risk information does not necessarily make them welfare-reducing, however; the true risks might justify the same choice. Indeed, as Carrillo and Mariotti (2000) discuss, people with inaccurately high estimates of the risks of particular behaviors, such as smoking, may remain “strategically ignorant” of the true risks in order to heighten their resolve to avoid those behaviors.
impact. In addition, Austin’s cake-eating behavior, or the weight gain that stems from it, may influence what other people eat.

The fact of failure may have costs of its own. As will be discussed below, rule violations can carry implications for an individual’s ability to resist future temptations. In addition, Austin may suffer guilt and other painful internal feelings as a result of having broken his “no cake” commitment (Loewenstein & O’Donoghue 2006); these emotions are independent of and in addition to the effects of the cake on Austin’s waistline, health, or future resolve. He may also be subjected to scorn, ridicule, or other shaming sanctions by those close to him, or may suffer from wider-ranging forms of societal disapproval based on cultural norms regarding healthy eating, slimness, and so on. While such informal penalties and norms are sometimes recommended to help people win their internal battles (see, e.g., Bloche 2005, 1350-54), magnifying the costs of failure makes any willpower lapses that do occur more damaging than they would otherwise be (see Loewenstein & O’Donoghue 2006; Glaeser 2006).

B. Exercise Costs

The concept of willpower implies the application of effort (Shefrin & Thaler 1988, 611). Accordingly, scholars have examined not only the failure costs associated with lapses, but also the costs of exerting willpower effort (whether successfully or not). Economic models of self-control often explicitly incorporate these costs (e.g., Gul &

---

13 This result is not inevitable; it depends on the insurer’s inability to appropriately price the added risk associated with Austin’s consumption choice. See Bhattacharya & Sood 2007.
14 Although the precise mechanism is unclear, a recent longitudinal study supports the idea that obesity can spread through social networks (Christakis & Fowler 2007). More generally, the social acceptability of obesity may change as the proportion of the population that is obese changes (ibid. 377; Philipson & Posner 2008, 976-77).
Pesendorfer 2001, 1420; Shefrin & Thaler 1988, 612; Ozdenoren et al. 2006). In addition, a growing body of experimental work confirms that willpower exertions constitute costly draws against cognitive resources. This appears to be true even in a very basic physiological sense; recent research indicates that the brain uses a great deal of glucose when engaging in self-control efforts (Gailliot et al. 2007; Gailliot & Baumeister 2007).

A series of experiments involving pairs of tasks undertaken in succession supports a short-run cognitive depletion effect (see, e.g., Baumeister & Vohs 2003). In the first of the paired tasks, subjects are required to exert self-control in some fashion not required of the control group, such as resisting a tempting plate of cookies (Baumeister et al. 1998), or suppressing thoughts (such as of a “white bear”) or emotional reactions (such as to sad or funny movies) (Muraven et al. 1998). The second task is frustrating or otherwise taxing, such as attempting to solve unsolvable puzzles or squeezing a handgrip. In each case, researchers found that participants who had been required to exercise self-control earlier performed less well or gave up sooner (Baumeister et al. 1998; Muraven et al. 1998; see Baumeister & Vohs 2003 (discussing these and other studies)). From this, researchers have concluded that willpower resembles a muscle that can become fatigued with use (e.g., Muraven & Baumeister 2000). More broadly, research suggests that people possess limited cognitive reserves that can become depleted by a variety of exertions, including those related to self-control (Baumeister et al. 1998). Some support for this broader proposition is found in research showing more indulgent snack choices among “impulsive” individuals following more difficult cognitive exertions (Shiv & Fedorikhin 1999), as well as in studies indicating that making choices can reduce the ability to exercise self-control (Vohs et al., 2008; see Baumeister 2003, 12-14 (reviewing studies on “decision fatigue”).
These studies provide important insights, but leave a number of questions open. For one thing, it is unclear how, and how frequently, one’s self-control reserves are replenished (Muraven & Baumeister 2000, 256) although research is beginning to explore the question (e.g., Tice et al. 2007). Other questions surround the conditions under which depletion occurs. For example, Dewitte et al. (2009) found that the types of tasks that follow each other in succession can make a difference, with self-control in one task improving rather than degrading subsequent performance of a similar task that draws on the same control processes. Also unclear is the degree to which self-control exertions, as distinct from the deprivations and frustrations associated with those exertions, deplete cognitive reserves. For example, the hungry subjects who were exposed to cookies but forbidden to eat them had to exercise self-control, but they also suffered the frustrating experience of being exposed to cookies that they could not have—and this latter effect could be produced independently of the former if cookies were presented in a manner that made them sensorily available but physically inaccessible.\(^{15}\) This is no mere detail, given that removing alternatives from individuals’ choice sets is often presented as a useful way to avoid both the costs of willpower exertions and the costs of willpower lapses.

In thinking about costly exertions of self-control, it is important to recognize that they can either be successful or unsuccessful in preventing a lapse. When the exercise is unsuccessful, the individual not only suffers the full measure of failure costs but also bears the costs of her failed efforts to stop the lapse. Even when the exercise is successful in stopping a particular lapse, the toll taken by the exercise may make later lapses (or other

---

\(^{15}\) Some studies not involving self control as such have suggested that volitional choice is the key to depletion. See, e.g., Baumeister et al. 1998 (effects of choosing to give a speech taking a particular position versus being assigned to do so).
cognitive shortfalls) more likely (e.g., Baumeister & Vohs 2003). If exercise costs are sufficiently high, stopping a lapse may not be efficient. Just as people may rationally choose not to incur costs to gather further information (Stigler 1961) or to further reduce the chance of errors (Demsetz 2008, 21), people may rationally choose not to exert willpower on a particular occasion. This analysis might seem to suggest that individuals should “choose their battles” and give in to temptations immediately when the expected costs of preventing a lapse are higher than the expected failure costs associated with the lapse (see, e.g., Loewenstein 2000b, 61; Ozdenoren et al., 2006; cf. Loewenstein & O’Donoghue 2006).

But giving in to temptation is not always a winning strategy, even if it seems cost-justified on a particular occasion. Exercise costs are dynamic: patterns of willpower exertions, successes, and failures can have implications for the cost and efficacy of future applications of willpower. For example, one of the most familiar self-control strategies is to adopt personal rules that intentionally “bundle” similar instances of particular decisions (e.g., Ainslie 2001, 153; Ainslie 1992, 142-73; Bénabou & Tirole 2004). This approach lowers exercise costs across a given domain by short-circuiting the internal deliberative exercises that might otherwise be triggered by each specific instance of temptation (see, e.g., Ainslie 2001, 112-13). By raising the stakes, however, such rules demand a higher expenditure of willpower at each choice juncture than would be cost-justified based on an examination of that specific choice in isolation (see, e.g., Ainslie 2001, 113-16).16 People can also successfully reduce exercise costs though self-distraction or attention-control

16 The rules can also inflict costs, however, both in depriving people of flexibility in instances where it would be welfare enhancing (e.g., Ainslie 2001, 143-60) and in amplifying the negative impacts of any lapse (e.g., Bénabou & Tirole 2004, 851-56). See infra Part V.C.
techniques (Mischel et al. 2003; Ainslie 1992, 133-35), or other kinds of precommitments (e.g., Schelling 1984; Ainslie 1992, 125-44; Elster 2000, 29-34; Ariely & Wertenbroch 2002); these strategies, too, will often have implications that extend beyond a given tempting episode.

C. Erosion Costs

Analysis of exercise costs (which come from exerting willpower) is incomplete unless we also take into account the costs that come from not exerting willpower—“erosion costs.” One subset of this category of costs has already been discussed: the effect of one lapse on the likelihood of later lapses. Bright-line rules are a powerful technology for fortifying self-control because breaking a personal rule can “set a precedent,” convey unwelcome information to the individual about herself, and harm her self-reputation (e.g., Schelling 1984, 74-75; Bénabou & Tirole 2003, 151-59; Bodner & Prelec 2003, 119-21; Ainslie 1985, 147). Yet, when a lapse occurs, these effects can make future exertions of willpower more difficult and less likely to succeed. Thus, Austin may conclude after eating a piece of cake that he has blown his diet (at least for the day) and that there is no additional harm in eating as much as he likes (Herman & Polivy 2003, 467-71). Similarly, a person who views a particular task as a “test of willpower” may find it easier to resist temptation because doing so conveys positive information about the strength of her own will (Magen & Gross 200, 417), but that same characterization of the task could also make failure more damaging (ibid., 427).

Erosion of willpower can also occur at a more general level, whether within one individual over time, or across society as a whole. Although characterized in a variety of
ways, these sorts of costs have received attention in both the theoretical and empirical literature. For example, Cooter (1991, 150) suggests that a “theory of the will” must include a “self-development” component. Likewise, Kuklin (1992, 666 & n.36) links John Stuart Mills’s “moral muscles” argument against paternalism to the contention that placing alternatives out of reach will deprive consumers of learning opportunities and the related strengthening of “moral fiber.” More recently (and at a higher level of generality) Klick and Mitchell (2006, 1626-27) have suggested that paternalistic policies present a variety of “cognitive hazards” and could “undercut personal incentives to invest in cognitive capital,” using self-control as one example (see ibid., 1631). Whitman and Rizzo (2007) have similarly emphasized the risks to willpower development of policy measures designed to make exercises of willpower unnecessary, including reduced opportunities for learning (ibid., 430-31) and a potential “crowding out” of internal control mechanisms (ibid., 432-33 (citing Fishbach & Trope 2005)).

Consistent with the muscle analogy, empirical work has found evidence of a “strength training” effect produced through repeated exertions of will (Muraven et al. 1999; Loewenstein 2000b, 56-57; Ozdenoren et al. 2006, 21-23). Muraven et al. (1999) conducted a longitudinal study that compared the performance of participants at two experimental sessions spaced two weeks apart, varying what the participants were assigned to do (if anything) during the intervening weeks. Although the study has some limitations, as the authors acknowledge, the results suggested that practicing some sorts of self-regulatory tasks, such as maintaining good posture, could improve one’s “stamina” and counter the depletion effects associated with exercising self-control (ibid. 452). Putting the findings of this study together with earlier findings of depletion effects, the authors
concluded that “[i]t is good to exert self-control on a regular basis because in the long run, these exercises will strengthen self-control and make a person less susceptible to the depleting effects of a single exertion.” (ibid., 456).

More broadly, individuals who engage in self-control (or who fail to do so) arguably contribute to cultural norms and to the intergenerational transmission of values supporting (or undercutting) willpower. Suggestive in this regard is research showing that different age cohorts exhibit different savings behaviors and monetary attitudes which do not seem to be fully explained by aging itself (see, e.g., Lunt & Livingstone 1992, 101-32; Daniel 1997, 180). The notion that having strong willpower is a valuable personal characteristic may be critical to mustering effort in response to challenges framed as tests of will (see Magen & Gross 2007, 426-27). All of these possibilities suggest that sometimes a short-run exercise cost comes with a longer-run benefit (cf. Klick & Mitchell 2006, 1625).

IV. Policy Responses and Rationales

Discussions of willpower and related problems of time inconsistency often reach starkly divergent conclusions about policy. These differences are driven in part by unresolved empirical questions about how willpower works, about the relative importance of failure costs, exercise costs, and erosion costs, and about the meaning and implications of willpower heterogeneity. Obviously, policy responses also depend on the operative normative objectives. The sections that follow work through a slate of policy alternatives that have received attention in the literature, including doing nothing.

A. Compensatory Transfers

The fact that willpower carries the potential to reduce welfare could have implications for redistributive policy, especially if low willpower correlates with poverty (see, e.g., Kaplow 2008, 361-62). There is some evidence that impatience is inversely related to cognitive ability (Dohmen et al. 2008; Benjamin et al. 2006)\(^{18}\) which in turn might correlate with income. There is also some work connecting wealth with self-control (Ameriks et al. 2003; Ameriks et al. 2007), although the contours of the relationship are not entirely straightforward (see Ozdenoren et al. 2006, 15-16). Significantly, self-control problems can manifest in both oversaving and undersaving (see, e.g., Ameriks 2003). The appropriate policy response would depend on a number of other issues as well, including the operative social welfare function (Shaviro 2007, 785) and whether well-being is measured on a lifetime or sublifetime basis (Adler 2007). The feasibility of transfers in favor of low-willpower individuals would also depend crucially on the extent to which willpower status is identifiable (cf. Weisbach 2007) and mutable.

Put into the terms introduced above, a compensatory approach would effectively buffer failure costs in each individual instance and, as a result, would also likely reduce the exercise costs that people choose to incur. Both of these results would be advantageous for those suffering from low willpower, but would be countered by two disadvantages. First, failure rates (and hence total failure costs) might rise as people do less to avoid lapses.

\(^{18}\) While impatience is not the same thing as low willpower, low willpower is one reason that impatient behaviors may at times be observed. For another take on the connection between intelligence and self-control, see Posner 1997, 28-29 (observing that “as imagination is a component of intelligence, a more intelligent person will be more future-oriented than will a less intelligent one” but also noting a countervailing factor—the intelligent person’s ability to “develop rationalizations that may deceive the future-oriented self”).
Second, reductions in willpower exertions over time could produce erosion costs that cash out in higher exercise and failure costs.

**B. Repricing Lapses**

A very different approach builds on the idea of Pigouvian taxation (Pigou 1932). In the interpersonal context, these taxes (or subsidies) seek to close the gap between the private payoff of an act and the act’s social payoff (ibid. 172-203). Similarly, taxes or subsidies might be used to make the current self’s calculus align more closely with the actor’s long-run interests (see Whitman 2006). “Sin taxes” are a well-known manifestation of this basic approach.19 Making the present self bear a higher cost to indulge her preferences could correct for internalities by shifting the costs of behavior to the actor choosing to engage in it.20 One complication is that the higher price can affect the decisions of everyone subject to it, not just those with low willpower who are enmeshed in internal battles (e.g., Whitman 2006). Of course, sin taxes can be (and often are) justified on grounds other than willpower augmentation, such as the externalities that conduct imposes on other parties, or even the desire to raise revenue. It is critical to disentangle the externality-based strand of the argument for repricing conduct through taxes and subsidies from the willpower-based strand of that argument, even though the two may interact at the level of policy (see, e.g., Strnad 2006).

19 Although my focus in this section is on pecuniary taxes or subsidies, repricing of conduct might also occur through the inculcation of responses like guilt or shame (see, e.g., Kaplow & Shavell 2007; Loewenstein & O’Donoghue 2006).

20 This only works as advertised if the present self actually has to bear the added cost. If credit or accumulated savings are available, the present self can offload both the harms associated with the behavior and the higher price of engaging in it onto her future self (see Whitman & Rizzo 2007, 428-29).
Suppose, for example, that a tax is imposed on conduct like smoking or eating fatty foods because of externalities. The tax may be sufficient to change an individual’s consumption choice, even if it ignores the possibility of intrapersonal struggles (Dodd 2008). In such a case, a fortuitous side effect of an externality-based repricing strategy could be making people better off by their own lights (by keeping them from doing what they did not “really” want to do in the first place). By the same token, repricing that corrects for internalities could have the happy side effect of altering decisions in ways that benefit other people. But a policy instrument that does not correct for all unaccounted-for costs (both internalities and externalities) relevant to a given actor’s decision will do less to reduce suboptimal decisionmaking. Moreover, justifying a repricing policy (in whole or in part) on the grounds that it usefully augments willpower depends on two empirical beliefs that are unlikely to be true across the board for a given population: first, that the long-run self has preferences that are in alignment with the policy promoted by the price change; and second, that honoring those long-run preferences will make the individual better off. To skip a step and simply base the repricing policy on the belief that it will make people better off is not a justification based on willpower problems (or what Calabresi and Melamed (1972) term “self-paternalism”) but rather a straightforward exercise of paternalism.

Another problem relates to heterogeneity in the responsiveness of individuals to price changes. If a low-willpower individual finds herself unable to resist a given temptation despite the added penalty, she must suffer not only the failure costs associated

\footnote{Just as externalities do not always cause people to choose differently than they would in the absence of the externality, but do tend to distort decisionmaking, so too with any unaccounted-for costs that a policy instrument does not reach. See note 7 \textit{supra}.}
with making a choice that makes her worse off, but also the costs of the penalty (e.g., Strnad 2005, 1254; Dodd 2008, 86). For example, someone who smokes must pay both the tax (now) and the costs in diminished health (later) (e.g., Strnad 2005, 1254). If the existence of the penalty made her struggle harder before succumbing to the temptation, racking up extra exercise costs in the process, so much the worse.23

C. Removing Choices

One apparent way out of the bind just described is simply to remove tempting choices (or force people to take the more patient or far-sighted choice). Choice removal can take a variety of forms and, like repricing, may be motivated by factors other than self-control concerns. The government might regulate product features directly (as by restricting loan pricing arrangements or permissible levels of nicotine) or control the time, place, or manner in which a particular product can be purchased or consumed. A product might be made available only to those who meet certain criteria (such as age limits), or only when bundled with another feature (such as a delay or a counseling session), or only when consumed independent of certain other features (such as on-site advertising).

If a welfare-reducing alternative is missing from the choice set, failure costs associated with taking that choice are eliminated. Likewise, because no exertion of will is necessary to avoid selecting an unavailable alternative, exercise costs are eliminated as

22 Note, however, that one response to this “double payment” problem would involve holding the fines in trust for the individual’s later self, or making the fines into a kind of forced insurance purchase (Strnad 2005, 1254). Pursuing this strategy would amount to a hybrid approach that is partly about removing choices, inasmuch as it would no longer be possible to purchase the desired product without also purchasing what amounts to an insurance package.

23 With exercise costs added into the mix along with penalty costs and the background failure costs, the penalty might actually be triple the failure costs, or perhaps even more than that if subjective feelings of guilt and shame are heaped on top (see Loewenstein & O’Donoghue 2006, 183).
well. For example, Kumru & Thanopoulos (2008, 774-75) suggest that savings in the costs of exerting self-control could form a rationale for Social Security. Such approaches raise some concerns, however. First, because of preference heterogeneity, removing choices across the board is likely to interfere with the realization of some people’s true, long-run preferences (see, e.g., Whitman 2006). A second and more fundamental concern relates to one of the very features that makes choice removal such an attractive strategy—its ability to render exercises of willpower unnecessary (see, e.g., Posner 1997, 32; Kumru & Thanopoulos 2008, 774-75). Although little is known about how a society’s stock of willpower (and willpower-fostering values and norms) is built up or depleted over time, there is at least the potential for erosion costs if people are not called upon to exercise self-control on a regular basis. One would need to compare any such costs with the failure costs and exercise costs that choice removal would eliminate. Finally, it is unclear whether coercively applied choice removal would be identical from an erosion costs perspective to choice removal at the behest of the individual, and whether the same effects obtain when an individual affirmatively chooses to deprive herself of a choice as when choice removal is made the default option.24

D. Facilitating Precommitments

Any policy that purports to address self-control problems must contend with a foundational problem that the preceding sections have emphasized: we cannot infer from

24 Default selections can be viewed as a form of repricing that makes the default choice cheaper in terms of hassle and attention requirements than the other choices. Even when the costs of opting out appear quite low, as in the “one-click paternalism” advocated by Thaler and Sunstein (2008, 248-49), the repricing still seems to have significant effects on behavior—at least to the extent that the choice to stick with the default stems from inertia rather than some informational content or “implicit advice” gleaned from the default selection (see Choi et al. 2003).
behavior alone whether people are suffering from willpower difficulties. Policies that tax or eliminate particular choices across the board are likely to be overinclusive, interfering with many consistently-held preferences. Voluntary precommitment mechanisms, in contrast, could induce people with known willpower problems to self-identify, better accommodating willpower heterogeneity (e.g., O’Donoghue & Rabin 2003a; Bhattacharya & Lakdawalla 2004). Not everyone with low self-control will recognize their own weaknesses, however, and even those who understand that they will confront willpower problems may underestimate the extent of the difficulties they will face (e.g., Rabin & O’Donoghue 2003a). Thus, a policy that relies on self-identification will likely be underinclusive. Nonetheless, precommitment programs have the advantage of avoiding intrusions on autonomy while selectively delivering benefits to those who desire them.

The value of precommitment to people who foresee encountering self-control problems has been recognized ever since Homer penned the tale of Ulysses and the Sirens (see, e.g., Strotz 1955; Elster 1979; Laibson 1997). An important question for legal policy is whether this demand for precommitment can be met without state involvement. People can and do privately precommit in all sorts of ways, from putting the alarm clock on the other side of the room to socking away money in illiquid assets (Laibson 1997). It is also possible for private entities to offer precommitment products. Thus, casinos may offer self-exclusion programs (Harford 2008, 61); instructors may enforce interim deadlines selected by students (Ariely & Wertenbroch 2002), and enterprises like the recently formed StickK.com can offer to channel money away from those who break their self-imposed commitments. Private precommitment mechanisms are likely to be underprovided, however. Not only may they add administrative costs, they may operate at cross purposes
with the profit motives of the private entity in question. Moreover, even when such products are offered, they may lack enforcement teeth (see, e.g., Schelling 1984, 102-03) or be prone to unraveling.

Governmentally sponsored precommitment products, in contrast, could be coercively enforced. While the government is already involved to some extent in delivering precommitment opportunities through state-sponsored self-exclusion policies for problem gamblers (see, e.g., Leitzel 2008) and through its treatment of certain savings vehicles (see, e.g., Weiss 1991, 1313-14), it could clearly do more (e.g., ibid.). A wide range of innovative approaches have been discussed in the literature, including allowing people to opt into particular tax or subsidy schemes (see, e.g., O’Donoghue & Rabin 2003b; Bhattacharya & Lakdawalla 2004; Fennell 2005, 1482-85). While the cost of providing precommitment opportunities would represent a form of redistribution in favor of those with low willpower (cf. Mitchell 2005), the relief would be targeted in nature and might be justified on the distributive grounds suggested above. A precommitment package could be analogized to other in-kind goods and services, such as those provided to people with disabilities, that offer benefits to the target population without inducing others to emulate that population (see Weisbach 2007).

---

25 This may help explain the dearth of credit cards that allow customers to set binding limits for their own spending (see, e.g., Ariely 2008, 124-26). There could be instances, however, in which a private interest in controlling externalities, such as an employer’s desire to improve the health of her employees to reduce absenteeism and control insurance premiums (see Henderson 2009), would dovetail with the provision of precommitment devices.

26 The potential for unraveling is clearest in financial contexts where the fact that funds are tied up in one place would not preclude their use as collateral to obtain funds elsewhere (see Laibson 1996, 27 & n. 13), although other kinds of precommitments might also be undermined (see, e.g., Gruber & Köszegi (2001, 1286).
Yet another issue involves the substantive content of the precommitment. We might expect the government to sponsor only programs targeted at myopic choices like overconsumption, leaving people who wish to avoid hyperopic problems like overwork (see Kivetz & Simonson, 2002; Hamermesh & Slemrod 2008) empty handed. In addition, the questions discussed above about which self’s preferences should trump reappear in this context, along with the concern that people would precommit to programs of conduct that would make them worse off. A particular concern is that people would bind themselves to overly austere regimens. The problems are similar to those that afflict people who try to enforce overly strict rules against themselves (e.g., Ainslie 2001, 155-56), except that the state may be relatively better positioned to actually move choices out of reach or, alternatively, coercively apply taxes to particular consumption choices. Meaningful enforcement would make harmful lapses from the prescribed program less likely but more painful, while excessive strictness could exact a heavy price in foregone utility.

E. Doing Nothing

Given the various countervailing considerations associated with each of the approaches above, a question inevitably arises: would we do better as a matter of legal policy to simply ignore willpower problems? Such a hands-off approach might flow from skepticism about governmental interference with free choice and the learning processes associated with it (see, e.g., Mitchell 2005; Whitman 2006; Klick and Mitchell 2006; Glaeser 2006).27 Another reason for caution is suggested by Besharov’s (2004) application

27 Some who oppose much of the “new paternalist” program still support measures, such as those designed to induce people to pay closer attention to their own behavior, on the ground that they enhance choice (e.g., Mitchell 2005, 1255-57).
of the theory of the second best to cognitive errors.\textsuperscript{28} Interventions to buttress self-control could unwittingly keep low willpower from offsetting other cognitive errors, such as a projection bias or overoptimism (see Besharov 2004; Whitman & Rizzo 2007, 427-28). For example, enforcing the stated long-run preferences of a person who takes on unrealistic projects or mispredicts her reactions to a particular change in circumstances may consign her to misery (see, e.g., Besharov 2004).

While these cautions should be borne in mind in shaping policy, doing nothing is in many ways an illusory option. As already suggested, even policies that are wholly motivated by externalities may bear on conduct that is also the subject of intrapersonal struggles. Where a choice exists about the precise means by which to pursue such policies, there no cogent reason to ignore willpower-related effects. For example, suppose that policymakers must decide whether to address a particular unwanted behavior by imposing a penalty at the time of the behavior or by imposing a present-value equivalent tax on the behavior at a later time. If there are many individuals with self-control problems who would like to avoid engaging in the behavior for their own reasons, the earlier penalty might do better at changing behavior because it will help to leverage people’s own preferences into actions. Indeed, even interspersing a delay between a particular consumption choice and its fulfillment could lead many people to comply with the socially optimal policy based on their own optimizing decisions (see e.g., Utset 2009)—although not without some cost (see Loewenstein & Haisley 2008, 239 n. 7). Such factors seem

\begin{flushleft}
\textsuperscript{28} The theory of the second best, initially formalized by Lipsey and Lancaster (1956), suggests that efforts to correct a single identified impediment to optimality could make matters worse, as where the problem under examination counteracts another, larger threat to optimality within the same system. See, e.g., Markovits 1998.
\end{flushleft}
relevant because they could reduce the degree to which the law must rely on coercion in order to achieve its ends.

Similarly, we might find the welfare impacts on low-willpower and high-willpower people, respectively, relevant to a choice about whether to respond to externalities by removing a given choice or merely repricing it. While repricing can better accommodate heterogeneity in preferences, it also presents risks for people with severe willpower problems to the extent it exposes them to a potential double penalty (e.g., Strnad 2005). In a related vein, tax policy must make choices about matters like how and when to tax people, and whether and how to tax savings—decisions that will inevitably have differential effects on people with different willpower levels (see, e.g., Fennell 2009). Thus, even if one believes that willpower problems do not make out an affirmative case for government intervention, self-control issues may still be policy relevant.

V. Directions for Future Research

Research at the intersection of willpower and legal policy is undergoing a tremendous surge of creativity and productivity. Several research issues have been flagged in the course of the discussion above, but three areas in need of further study seem especially important to mention in closing.

A. Unpacking Heterogeneity

While a great deal of work has gone into understanding how willpower operates, researchers are only beginning to understand individual differences in self-control. Some existing empirical work already attempts to measure willpower and determine its
correlates, but additional research is necessary to understand the extent to which self-control levels respond to particular inputs or remain entrenched over time. A promising avenue is suggested by recent attempts to break apart willpower capacity from willingness to exert self-control effort (Tittle et al. 2004). Similar in spirit is work that examines how background or “trait” self-control impacts the rate of depletion associated with willpower exertions (see Muraven et al. 2005) or the effects on willpower of other cognitive burdens (Shiv & Fedorikhin 1999). Also important to untangle is the role of tempting opportunities in producing differences in willpower-related outcomes independent of differences in background willpower levels. For example, a person who has no ability to borrow will not face the same temptations to spend as someone with an open-ended line of credit (see, e.g., Bar-Gill 2004; cf. Laibson 1997; Tittle et al. 2004, 149-50). Similarly, if a person gets little utility out of short-run indulgences, passing them up requires no great feat of will. In such cases, the selves simply have less to bargain over.

In addition, more attention should be given to heterogeneity among types of self-control exertions. Research on self-control has largely focused on rather contrived willpower exertions, such as not eating a cookie or suppressing a particular emotion or thought (see, e.g., Baumeister & Vohs 2003). For obvious ethical reasons, dangerous or illegal temptations are not employed (see Muraven et al. 2006, 274), and the constraints inherent in the experimental setting have limited the degree to which observed willpower dilemmas mimic those with which people chronically struggle. Future research should be directed at determining the extent to which different kinds of willpower exertions have

---

29 New technologies may help in this regard, however. For example, Muraven et al. (2005) equipped research participants with Palm Pilots that enabled them to record information in real time about matters such as mood, self-control demands, and drinking patterns.
different effects in terms of depletion, replenishment, and long-term strength training effects.

**B. Understanding Willpower Erosion and Spillovers**

Future research should also be directed to a better understanding of the interaction between self-control opportunities, the strength of the willpower response, and the cultural transmission of willpower. Regardless of whether willpower issues justify a policy intervention, policy choices are likely to have effects on the costs associated with self-control problems. External constraints on behavior, such as the removal of certain overly tempting consumption alternatives, can produce immediate gains for all those for whom resisting the temptation would be both desirable and costly. Yet longer-run effects on self-control must also be examined; if willpower erodes as result of its nonuse, members of a society with a diminished choice set may become increasingly unable to effectively regulate themselves.

A related question involves the extent to which willpower facilitates conduct that is socially valuable on balance (or, put differently, the extent to which willpower suppresses socially harmful conduct). To the extent that such a correlation exists, strengthening willpower society-wide may not only make individuals better off by their own lights, it may also help them to behave voluntarily in ways that are better for others.

**C. Delivering Bounded Exceptions**

Another unexplored role for policy is in offering bounded exceptions to limitations that people impose on themselves (whether on their own or through public or private
precommitment mechanisms). Very simple bright-line rules are easier to enforce against oneself than either complicated rules that require keeping track of amounts or instances (see Gilboa & Gilboa-Schechtman 2003, 132-34) or standards that call for judgment by each of one’s (unreliable) present selves. But these rules are likely to overshoot the mark and prescribe excessive restraint (see, e.g., Ainslie 2001, 143-60). Rules that bind behavior above and beyond the optimal level create obvious pressures to make exceptions, but the very “bundling” strategy that makes the rule effective (ibid., 153) also limits resort to exceptions, lest they swallow the rule (see, e.g., Ainslie 1985, 148-49).

The results can be enormously costly. Not only may people suffer utility losses from complying with an overly strict regimen, an excessively restricted pattern of behavior may eventually sap self-control and produce lapses, which then produce further lapses. Ainslie (2001, 155-56) describes that problem this way: “After the need for clarity has taken its toll on subtlety, and overcaution has reduced flexibility, and misplaced caution with its consequent lapses has eroded resolve and corrupted self-observation – in short, after the makeshift nature of our attempt at deciding according to principle has caught up with us – our wills may wind up getting in the way of our longest-range interests.”

A third party capable of both extending an exception to a general rule and appropriately firewalling it off from other similar instances would be able to offer an extremely valuable service. While a private party could potentially play this role, the government might be well-positioned to offer distinctive, bounded exceptions, at least in the financial realm. Indeed, the possibility that income tax refunds may (unwittingly) serve just such a function by offering a singular, infrequent source of “found money” has received some attention (e.g. Fennell 2006, 151). Future work examining the potential for
governmentally sponsored precommitment opportunities should not neglect the possibility that people may at times need not only ways to bind themselves, but also ways to give themselves a break.

**Literature Cited**


http://lsr.nellco.org/upenn/wps/papers/169


Baumeister RF. 2003. The psychology of irrationality: Why people make foolish, self-defeating choices. See Brocas & Carrillo 2003, pp. 3-16


Baumeister RF, Vohs KD. 2003. Willpower, choice, and self-control. See Loewenstein et al. 2003b, pp. 201-16


Fennell LA. 2009 Taxing willpower. Unpublished manuscript on file with author.


See Brocas & Carrillo 2003, pp. 127-36


Hamermesh D, Slemrod J. 2008. The economics of workaholism: We should not have worked on this paper. *B.E. J. of Econ. Analysis & Pol.*, 8: Article 3.

http://www.bepress.com/bejeap/vol8/iss1/art3


Johnson EJ. 2008. Man, my brain is tired: Linking depletion and cognitive effort in choice.

*J. Consum. Psych.* 18: 14-16


Manuck SB, Flory JD, Muldoon MF, Ferrell RE. 2003. A neurobiology of intertemporal choice. See Loewenstein et al. 2003b, pp. 130-72


http://www.nber.org/papers/w12278


Readers with comments should address them to:

Lee Anne Fennell
University of Chicago Law School
1111 East 60th Street
Chicago, IL  60637
lfennell@uchicago.edu
For a listing of papers 1–399 please go to Working Papers at http://www.law.uchicago.edu/lawecon/index.html

400. Shyam Balganesh, Foreseeability and Copyright Incentives (April 2008)
401. Cass R. Sunstein and Reid Hastie, Four Failures of Deliberating Groups (April 2008)
407. Cass R. Sunstein, Two Conceptions of Irreversible Environmental Harm (Mary 2008)
408. Richard A. Epstein, Public Use in a Post-Kelo World (June 2008)
410. Adam B. Cox and Thomas J. Miles, Documenting Discrimination? (June 2008)
413. Thomas J. Miles and Cass R. Sunstein, Depoliticizing Administrative Law (June 2008)
414. Randal C. Picker, Competition and Privacy in Web 2.0 and the Cloud (June 2008)
417. Omri Ben-Shahar, How to Repair Unconscionable Contracts (July 2008)
418. Richard A. Epstein and David A. Hyman, Controlling the Costs of Medical Care: A Dose of Deregulation (July 2008)
428. Irina D. Manta, Privatizing Trademarks (abstract only) (September 2008)
432. Cass R. Sunstein, Beyond Judicial Minimalism (September 2008)
433. Bernard E. Harcourt, Neoliberal Penalty: The Birth of Natural Order, the Illusion of Free Markets (September 2008)
437. Richard H. McAdams, Beyond the Prisoners’ Dilemma: Coordination, Game Theory and the Law (October 2008)
438. Dhammika Dharamapala, Nuno Garoupa, and Richard H. McAdams, Belief in a Just World, Blaming the Victim, and Hate Crime Statutes (October 2008)
443. Lee Anne Fennell, Adjusting Alienability (November 2008)
444. Nuno Garoupa and Tom Ginsburg, Guarding the Guardinas: Judicial Councils and Judicial Independence (November 2008)
446. Cass R. Sunstein and Richard Zeckhauser, Overreaction to Fearsome Risks (December 2008)
448. David A. Weisbach, Responsibility for Climate Change, by the Numbers (January 2009)
452. Richard Epstein, The Case against the Employee Free Choice Act (January 2009)
453. Adam B. Cox, Immigration Law’s Organizing Principles (February 2009)
454. Philip J. Cook, Jens Ludwig, and Adam M. Samaha, Gun Control after Heller: Threats and Sideshows from a Social Welfare Perspective (February 2009)
455. Lior Jacob Strahilevitz, The Right to Abandon (February 2009)
457. Lee Anne Fennell, Commons, Anticommons, Semicommons (February 2009)
464. Anupam Chander, Corporate Law’s Distributive Design (June 2009)
465. Anupam Chander, Trade 2.0 (June 2009)
467. Eric A Posner, Kathryn Spier, and Adrian Vermeule, Divide and Conquer (June 2009)
468. John Bronsteen, Christopher J. Buccafusco, and Jonathan S. Masur, Welfare as Happiness (June 2009)
469. Richard A. Epstein and Amanda M. Rose, The Regulation of Sovereign Wealth Funds: The Virtues of Going Slow (June 2009)
470. Douglas G. Baird and Robert K. Rasmussen, Anti-Bankruptcy (June 2009)
472. Bernard E. Harcourt, Neoliberal Penalty: A Brief Genealogy (June 2009)
473. Lee Anne Fennell, Willpower and Legal Policy (June 2009)