NBER WORKING PAPER SERIES

BEHAVIORAL PUBLIC ECONOMICS: WELFARE AND POLICY ANALYSIS WITH NON-STANDARD DECISION MAKERS

B. Douglas Bernheim
Antonio Rangel

Working Paper 11518
http://www.nber.org/papers/w11518

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2005

This paper is forthcoming in Economic Institutions and Behavioral Economics, edited by Peter Diamond and Hannu Vartiainen. We would like to thank Colin Camerer, Peter Diamond, Emmanuel Saez, and Nick Stern for useful comments. Antonio Rangel gratefully acknowledges financial support from the NSF (SES-0134618) and SIEPR. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers
B. Douglas Bernheim and Antonio Rangel
NBER Working Paper No. 11518
July 2005
JEL No. D0, D1, D6, D9, H0, H1, H4

ABSTRACT

This paper has two goals. First, we discuss several emerging approaches to applied welfare analysis under non-standard ("behavioral") assumptions concerning consumer choice. This provides a foundation for Behavioral Public Economics. Second, we illustrate applications of these approaches by surveying behavioral studies of policy problems involving saving, addiction, and public goods. We argue that the literature on behavioral public economics, though in its infancy, has already fundamentally changed our understanding of public policy in each of these domains.

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
bernheim@stanford.edu

Antonio Rangel
Department of Economics
Stanford University
Stanford, CA 94305
and NBER
rangel@stanford.edu

1. Introduction

Public economics has positive and normative objectives; it aims both to describe the effects of public policies and to evaluate them. This agenda requires us to formulate models of human decision-making with two components – one describing choices, and the other describing well-being. Using the first component, we can forecast the effects of policy reforms on individuals' actions, as well as on prices and allocations. Using the second component, we can determine whether these changes benefit consumers or harm them.

Traditionally, economists have made no distinction between the behavioral and welfare components of economic models. Such a distinction has not been necessary because standard welfare analysis is grounded in the doctrine of revealed preference. That is, we infer what people want from what they choose. When evaluating policies, we attempt to act as each individual's proxy, extrapolating his or her likely policy choices from observed consumption choices in related situations.

Interest in behavioral economics has grown in recent years, stimulated largely by accumulating evidence that the standard model of consumer decision-making provides an inadequate positive description of human behavior. Scholars have begun to propose alternative models that incorporate insights from psychology and neuroscience. Some of the pertinent literature focuses on behaviors commonly considered "dysfunctional," such as addiction, obesity, risky sexual behavior, and crime. However, there is also considerable interest in alternative approaches to more standard economic problems involving, for example, saving, risk-taking, and charitable contributions.

Behavioral economists have proposed a variety of models that raise difficult issues concerning welfare evaluation. No consensus concerning appropriate standards and criteria has yet emerged. Broadly speaking, there are two main schools of thought.

One school of thought insists on strict adherence to the doctrine of revealed preference for the purpose of economic policy evaluation. In this view, observed "anomalies" should be explained, when possible, by expanding the preference domain. Indeed, in the view of some economists, the only legitimate objective of behavioral economics is to identify preferences that robustly rationalize choices (Gul and Pesendorfer [2001,2004a,b]). This perspective maintains the tight correspondence between the behavioral and welfare components of economic models.

A second school of thought holds that behavioral economics can in principle justify modifying, relaxing, or even jettisoning the principle of revealed preference for the purpose of

welfare analysis. A number of possibilities have been explored. If people make systematic mistakes in identifiable circumstances, it may be appropriate to apply the principle of revealed preference *selectively* rather than systematically. If an individual's choices reveal several distinct sets of mutually inconsistent preferences, then normative evaluation may require the adoption of a particular perspective. If choices do not reveal coherent preferences, then perhaps normative evaluations should emphasize other aspects of well-being, such as opportunities. To pursue any of these possibilities, one must formulate separate, and potentially divergent, positive and normative models.

Adopting alternatives to the principle of revealed preference allows economists to engage on issues that specialists in other fields, as well as the public at large, regard as central policy concerns. For example, they can meaningfully address the "self-destructive" behavior of addicts or make sense of the claim that American's save "too little" for retirement.

However, there is also a danger. Revealed preference is an attractive political principle because it guards against abuse (albeit quite imperfectly in practice). Once we relax this doctrine, we potentially legitimize government condemnation of almost any chosen lifestyle on the grounds that it is contrary to a "natural" welfare criterion reflecting the individual's "true" interests. If we can classify, say, the consumption of an addictive substance as contrary to an individual's interests, what about choices involving literature, religion, or sexual orientation? If choices do not unambiguously reveal an individual's notions of good and bad, then "true preferences" become the subject of debate, and every "beneficial" restriction of personal choice becomes fair game.[2]

Given these dangers, if we are to relax the principle of revealed preference when evaluating public policy, it behooves us to set a high scientific threshold for reaching a determination, based on objective evidence, that a given problem calls for divergent positive and normative models. It is important to emphasize that any justification for modifying or replacing the principle of revealed preference must necessarily appeal to evidence other than observations of choice. After all, in the absence of additional assumptions, it is impossible to disprove the hypothesis that people prefer what they choose simply by examining their choices. As we argue in detail below, this is one respect in which direct evidence on the neural mechanisms of decision-making is beginning to prove valuable.

Unfortunately, behavioral economists have typically been somewhat cavalier in adopting normative criteria. For example, in the literature on quasi-hyperbolic discounting, it is now

[2] McCaffrey and Slemrod [2005] make a similar argument.

standard practice to adopt the "long-run" perspective ($\beta = 1$) for welfare analysis, rather than the perspective that governs "short-run" choices ($\beta < 1$). This approach has been criticized on the grounds that, according to the principle of revealed preference, the short-run perspective also has status as a welfare criterion. The arguments that have been offered in defense of the "long-run" perspective have not convinced skeptics that it is appropriate to attach absolutely no normative significance to short-run preferences. The foundations for welfare analysis therefore require closer attention.

This paper has two goals. First, we discuss emerging methods for normative policy analysis in behavioral economics, as well as potentially fruitful lines of inquiry. We explicitly argue against the view that any departure from the doctrine of revealed preference renders welfare analysis either infeasible or entirely subjective. Instead, we argue that it is sometimes possible to replace revealed preference with other compelling normative principles. For example, if one knows enough about the nature of decision-making malfunctions, it may be possible to recover tastes by relying on a *selective* application of the revealed preference principle. Accordingly, practicing behavioral economics requires us to modify – not to abandon – the key methodological principles of modern economics (see Rabin [2002] for a related argument).

Second, we review a collection of applications of behavioral economics to the field of public economics. In preparing this selective review, we have intentionally favored depth over breadth in the hope of providing a substantive discussion of welfare issues and policy implications. We focus on three specific policy issues: saving, addiction, and public goods. While each literature is still in its infancy, we argue that behavioral economics has already provided fundamental insights concerning public policy in each of these domains.

The remainder of this paper is organized as follows. Section 2 discusses alternative approaches to the problem of welfare. This section is an abbreviated version of Bernheim and Rangel [2005a], to which we refer the reader for additional details. Section 3, 4 and 5 survey applications to, respectively, saving, addictive substances, and public goods. Section 6 provides a brief discussion the future of behavioral public economics.

2. Conceptualizing and Measuring Welfare

Welfare analysis has two main components. First, one determines how policies affect the well-being of each individual. Second, one aggregates across individuals. As is well-known, the second step involves some thorny issues (e.g., those raised by Arrow's Impossibility Theorem).

However, since these are common to both neoclassical and behavioral approaches, we will say no more about them. Instead, we will focus on the assessment of each individual's well-being.

There is widespread agreement that normative criteria should respect the principle of individual sovereignty, which holds that notions of good and bad for society should be rooted in the notions of good and bad held by the affected individuals. This principle instructs policy analysts to act as each individual's proxy when comparing alternative policies. It precludes the analyst from imposing his or her own value judgments. Our focus here is, in effect, on the meaning of the phrase, "acting as each individual's proxy."

In the neoclassical paradigm, the analyst attempts to determine which policy choice the individual would make, given the opportunity. This is obviously difficult, since the policy choices under consideration differ considerably from the private choices that people ordinarily make. The beauty and power of standard consumer theory resides in the fact that it allows us to extrapolate choices among public policy outcomes from observations of private choices.

One common interpretation of the neoclassical approach is that people have well-defined preference rankings which the analyst discovers by examining evidence on choices (through the principle of revealed preference). These rankings are then taken as the basis for welfare evaluations. As detailed in Bernheim and Rangel [2005a], this interpretation rests on the following four assumptions.

Assumption 1: Coherent preferences. Each individual has coherent, well-behaved preferences.

Assumption 2: Preference domain. The domain of each individual's preference rankings is the set of lifetime state-contingent consumption paths.

Assumption 3: Fixed lifetime preferences. Each individual's ranking of lifetime state-contingent consumption paths remains constant across time and states of nature.

Assumption 4: No mistakes. Each individual always selects the most preferred alternative from the feasible set.

It is important to emphasize that the third assumption does not rule out the possibility that tastes vary over time or across states of nature. To illustrate, consider the following problem: choose either an immediate five-day vacation, or a ten-day vacation after a three month delay. The third assumption allows for the possibility that the preferred choice changes with age, or fluctuates randomly with mood. For example, if an individual is under stress, the immediate vacation may be more attractive. The assumption does *not*, however, allow for the possibility that, while in a relaxed mood, the individual would wish to prescribe for himself a different choice than he would actually make at other points in time while in an stressed mood. On the contrary,

while in a relaxed mood, he should regard the decisions he makes at other points in time while in stressed moods as optimal. Though he is willing to make different tradeoffs at different points in time and in different states of nature, his notion of a "life well-lived" remains fixed.

Another interpretation of the neoclassical approach, discussed at greater length in Bernheim and Rangel [2005a], holds that revealed preferences are merely constructs for systematizing information concerning choices. This view does not require one to take a position as to whether people actually have preferences, or whether revealed preferences coincide with "true" preferences. Rather, it posits that people act *as if* they optimize given particular preferences, and uses this representation to extrapolate choices among policy alternatives. According to this view, the neoclassical paradigm is only about choice.

Throughout the remainder of this section, we adopt the perspective that preferences are "real" objects. In our view, the concept of preference is something that we all understand in concrete terms. Even if we are limited to inferring others' preferences from their choices, this does not call the existence of preferences into question. After all, most of us believe we can learn much about our own preferences from introspection. None of us have ever chosen between spending two weeks on Maui and two years in prison, yet we know we would be happier with the first alternative; we do not need to infer this preference from an actual choice. From this perspective, the discovery of true preferences is a central objective of welfare economics.

One can think of the various approaches to welfare analysis that have appeared in the behavioral literature as efforts to grapple with the distinctive issues that arise when we relax each of the four assumptions listed above. We will consider each of them in turn.

2.A. Relaxing the first assumption (coherent preferences)

The first assumption holds that people have well-defined, coherent preferences. If observed choices are highly context-dependent, with significant decisions turning on minor and seemingly irrelevant aspects of framing (see, e.g., Tversky and Kahneman [1986]), then it may be appropriate to assume that people have poorly behaved or incoherent preferences (or possibly no preferences at all). In this case, how does one evaluate an individual's well-being?

One possibility is to abandon the principle that the welfare criterion used to evaluate public policy should be based on individual notions of good and bad *allocations*. Unlike the standard approach, this leads to a sharp separation between positive models describing choice, and normative models describing welfare. One interesting example of this approach appears in Sugden [2004], who argues for a notion of welfare based on *opportunities*. Sugden formulates a rigorous welfare criterion along these lines, and proves a counterpart to the first welfare theorem.

There are many practical and philosophical reasons to consider welfare standards based on opportunities rather than allocations (see, e.g., Cohen [1989], Sen [1992], and Roemer [1998]). This certainly simplifies some aspects of measurement, and it avoids the need to systematize behavioral observations by imposing untested assumptions. Yet we suspect that most economists will resist such a radical departure from the standard approach. Even if we acknowledge that opportunities are important, people also appear to care a great deal about allocations and subjective perceptions of well-being. And while there is *some* evidence of context-dependence and incoherence, we doubt anyone would claim that preferences are *entirely* incoherent (e.g., one can't induce the typical person to exchange two weeks at a resort in Maui for two years in prison by manipulating framing). An approach based *exclusively* on opportunities would appear to ignore this potentially valuable information.

2.B. Relaxing the second assumption (preference domain)

Some behavioral anomalies that defy explanation within the standard approach may become explicable if we expand the preference domain. Conceptually, this permits us to conduct welfare analysis by applying the principle of revealed preference, as in the standard approach (that is, we can use essentially the same model to describe choices and welfare). We discuss two examples.

The first example involves temptation and self-control. Motivating behavioral anomalies include evidence of apparent time-inconsistency and various forms of precommitment. Gul and Pesendorfer [2001] argue that it is possible to account for a range of otherwise puzzling behavioral observations if preferences are defined over both allocations and *choice sets* (see also Gul and Pesendorfer [2004a,b]). If some choices feel tempting when they are available, and if this detracts from well-being, then an individual may prefer small choice sets to large ones. This provides a reason to constrain future alternatives *even when constraints have no impact on choices*. In the Gul-Pesendorfer framework, a desire to constrain future choices does not imply that preferences change over time. On the contrary, as in the standard framework, the individual applies the same set of lifetime preferences at every moment in time. Even though, at time t, he might wish to constrain his available options for time $s > t$, he nevertheless approves of the choice he would actually make at time s in the absence of this constraint (because he understands the significance of temptation). In this framework, if one imposes suitable structure on choice data, one can discover lifetime preferences over allocations and choice sets by applying the principle of revealed preference, and one can use these preferences to make welfare evaluations, just as in the standard approach.

The second example involves social preferences. Motivating behavioral anomalies include, among others, a tendency to give money away in settings where there is no room for reciprocity (see, e.g., Camerer [2003] for a review of evidence on the dictator game), an apparent aversion to inequality (e.g., Fehr and Schmidt [1999] and Bolton and Ockenfels [2000]), and a desire to conform to group norms (see Jones [1984] for a review of pertinent evidence). For the purpose of positive modeling, behavioral economists frequently assume that preferences are defined not only over an individual's own consumption bundle, but also over social outcomes, such as the consumption bundles of others. If one imposes suitable structure on choice data, one can once again discover these tastes by applying the principle of revealed preference. These preferences provide a foundation for normative evaluation (in other words, one again uses essentially the same model to describe choices and welfare).

2.C. Relaxing the third assumption (fixed lifetime preferences)

The third assumption states that preferences over lifetime state-contingent consumption paths do not change over time or across states of nature. Behavioral anomalies motivating relaxation of this assumption include, again, evidence of apparent time-inconsistency and various forms of precommitment. From a positive perspective, a common modeling strategy involves endowing the individual with different well-behaved lifetime preferences at different points in time (Laibson [1997], O'Donoghue and Rabin [1999b, 2001]); one could, of course, also allow lifetime preferences to vary across states of nature (Loewenstein [1996], Loewenstein and O'Donoghue [2004]). Assuming we've properly measured these preferences, welfare analysis requires us, in effect, to adjudicate conflicts among them. The problem is analogous to welfare aggregation involving many individuals; here, we aggregate over multiple "selves."

One branch of the literature exploits this analogy. Effectively, it envisions person A at time t as the "child" of person A at time $t-1$. It then applies standard multi-person welfare principles. One possibility is to apply the Pareto criterion (see, e.g., Phelps and Pollack [1968], or Laibson [1997] and Battacharya and Lakdawalla [2004] for recent examples). The main problem with this approach is that the criterion is not very discerning. As a result, it is often impossible to rank interesting classes of policies. One usually ends up being able to offer policy makers little in the way of clear guidance. A second possibility is to aggregate preferences through the application of some welfare function. As in problems with multiple consumers, one can write down a class of well-behaved aggregators (i.e., the analog of Samuelson-Bergson social welfare functions) and attempt to derive general results. However, unless one has a basis for making specific assumptions about the aggregator, this approach fails to sharpen the prescriptions

generated from application of the Pareto criterion. Alternatively, one could in principle provide the policy maker with a mapping from properties of the aggregator (e.g., welfare weights) to prescriptions.

A second branch of the literature makes welfare evaluations based on some reasonably stable component of preferences. For example, O'Donoghue and Rabin [1999b] argue for the application of a "long-run" welfare criterion ($\beta = 1$) in models with quasi-hyperbolic discounting. In Bernheim and Rangel [2005a], we provide a formal justification of this criterion based on aggregation principles. In particular, we demonstrate that if the consumer's horizon is sufficiently long, and if the policy analyst applies any member of a large class of well-behaved aggregators, the resulting welfare criterion is "close" to long-run preferences. The intuition for this result is that the consumer judges tradeoffs between period t and $t + 1$ by exactly the same criteria in all periods but one, and the influence of any one "self" must decline to zero as the number of selves becomes large.

One can make a similar point concerning states of nature. To illustrate, consider an individual who lives in continuous time. Choices are essentially instantaneous but have long-lasting consequences (as an example, think of drug use). The individual's mental state is either "cold," which corresponds to one set of lifetime preferences, or "hot," which corresponds to another. Normally, the individual operates in a cold mode. At each moment, there's some chance that he enters the hot state, which has a fixed duration of ε. Suppose we model the arrival of the hot state as a failure-time process, with a fixed hazard parameter. As ε approaches zero, the fraction of time spent in the cold state converges to unity. Accordingly, if we aggregate preferences according to the frequency with which they prevail, we end up using the cold preferences for normative analysis. Even so, cold preferences do not describe behavior in this limit. Since hot states can create "momentary lapses" with long-lasting effects, the appropriate positive and normative models diverge. See Bernheim and Rangel [2005a] for a formal treatment.

2.D. Relaxing the fourth assumption (no mistakes)

The fourth assumption holds that choice and preferences do not diverge. Gul and Pesendorfer [2002] defend this assumption as follows: "Revealed preference theory defines the interest of people to be what they do. Since there is no objective standard of self-interested behavior it is unclear what it would mean for an agent to act against his self-interest."

Yet there are clearly situations where virtually everyone would agree that divergence does occur – where a choice is obviously not in someone's interest. There are also situations in which most would agree that public policy should recognize these divergences.

Consider the following example. American visitors in London suffer numerous injuries and fatalities because they often look only to the left before stepping into streets, even though they know traffic approaches from the right. This is a systematic pattern; one can't dismiss it as an isolated incident. A literal application of the revealed preference compels us to conclude either that these people simply have a very strong preference look left, or that they're masochistic. If we use these revealed preferences for welfare analysis, there's no legitimate basis for preventing someone from stepping in front of a truck. And yet, it's safe to say that, after recognizing the purpose of the intervention, anyone would be grateful. The pedestrian's objective -- to cross the street safely -- is clear, and the decision is plainly a mistake.

As another example, consider the treatment of children. Few economists would apply notions of consumer sovereignty and revealed preference to evaluate the welfare of a child. We acknowledge that children do not know what's best, and that their actions often fail to reflect valid preferences, probably because they give insufficient weight to consequences. Policies prohibiting the sale of cigarettes and alcohol to minors are therefore relatively uncontroversial. And yet, it's difficult to justify, objectively, the sense in which the revealed preferences of an irresponsible nineteen-year-old are legitimate, whereas those of a fourteen-year-old are not. While turning eighteen has profound legal significance, it doesn't discontinuously change the mechanics of decision-making.

There are other contexts for which revealed preference seems untenable as a guiding principle for public policy evaluation. For example, when people have sufficiently severe diagnosed psychiatric disorders, the state can and should step in to protect them. Eating disorders, while not quite as extreme, provide another illustration. For the purpose of public policy, we probably should not proceed on the assumption that an anorexic's refusal to eat is just an expression of valid preferences. On the contrary, we should and generally do regard this as dysfunctional. These examples are instructive because they suggest that, in some circumstances, it is reasonable to use evidence of brain process malfunctions – something other than choice data – to trump the principle of revealed preference. In these situations, denying the possibility of mistakes while rigidly adhering to the principle of revealed preference guarantees the use of an improper welfare criterion.

So far, we have confined our discussion to “dysfunctional” choices. More generally, almost any behavioral anomaly motivating some relaxation of the first three assumptions can also

motivate relaxation of the fourth. For example, evidence of time-inconsistent present-bias may reflect a systematic tendency to "over-consume." Likewise, people may make precommitments to prevent themselves from repeating a pattern of mistakes.

A natural analytic strategy involves endowing the individual with well-behaved lifetime preferences, while simultaneously specifying a decision process (or decision criterion) that does not necessarily involve selecting the maximal element in the preference ordering. To conduct positive analysis, one employs a model of the decision process (or criterion). To conduct normative analysis, one uses a model of lifetime preferences. In contrast to the standard approach, these positive and normative models potentially diverge.

Our model of addiction (Bernheim and Rangel [2004]), discussed in greater detail below, exemplifies this approach. We assume that people attempt to optimize given their true preferences, but randomly encounter conditions that trigger systematic mistakes, the likelihood of which evolves with previous substance use. One can also interpret the familiar model of quasi-hyperbolic discounting along similar lines (indeed, many of those who advocate this model favor this interpretation). In this interpretation, present-biased behavior is a mistake that results from the decision making processes' tendency to place too much weight on immediate rewards relative to future rewards.[3]

In justifying and implementing this approach, we encounter two critical and difficult issues. First, how do we know that choices and preferences diverge? That is, what is the basis for overturning the principle of revealed preference? Second, if we find compelling evidence of divergence, how do we identify preferences empirically? Both questions are addressed in the literature, though not in a single paper.

1. Criteria for overturning revealed preferences. With respect to the first issue, it is important to acknowledge that, strictly speaking, it is impossible to overturn the principle of revealed preference using only observations of choices. While choice experiments can overturn specific structural assumptions, overturning the principle itself necessarily requires other types of evidence. It is always possible to rationalize choice data by assuming that tastes are sufficiently context-specific.

One promising approach is to use evidence from neuroscience and psychology on the neural processes at work in decision making. For example, if it is possible to isolate a process that provides *inputs* for decision-making, and to show that this process either has substantive limitations, or that it malfunctions under identifiable circumstances, then the evidence may provide a foundation both for asserting the existence of errors, and for a particular reduced-form

[3] McClure et. al. [2004] present evidence that potentially supports this interpretation.

model of the error-producing mechanism. In this regard, brain processes of particular interest include those involved in anticipating and evaluating the outcomes of different choices, remembering pertinent information (memory), and attending to relevant data and options (attention). An example of this approach appears in Bernheim and Rangel [2004], where we argue that addictive substances interfere with the proper operation of an automatic neural forecasting system, thereby skewing decisions. We elaborate on this example in Section 4.E, below.

2. Strategies for identifying preferences. With respect to the second issue, it may be possible in a given instance to identify preferences by interpreting the available data through the lens of structural modeling. This approach requires one to formulate two tightly parameterized models – one for preferences, and one for choices. Ideally, it should be possible to justify the major structural assumptions of the decision-making model through the type of neurological and psychological evidence used to establish the existence of a discrepancy between preferences and choices.

As long as true preferences *influence* choices, even if the individual does not optimize, there will be some relationship between the parameters of the positive and normative models, and this will be useful for purposes of identification. Indeed, for the two examples mentioned so far (stochastic mistakes, as in Bernheim and Rangel [2004, 2005b], and quasi-hyperbolic discounting, as in Laibson [1997] and O'Donoghue and Rabin [1999b, 2001]), the parameters of the normative model are a *subset* of the parameters of the positive model (certain parameters describe true preferences, and others describe discrepancies between choices and preferences). Consequently, by estimating the positive model, one can recover preferences under the maintained hypothesis that the structural assumptions are correct.

Ideally, the assumed structure should subsume the possibility that there is no discrepancy between preferences and choices, so that it is possible to test this hypothesis. Both of the examples considered above satisfy this requirement.

Identification of preferences through choice data. As long as the parameters of the normative model are a subset of the parameters of the positive model, one can in principle estimate these parameters using data on choices, and nothing else. For example, Laibson et. al. [2004] use consumption data to parameterize a model with quasi-hyperbolic discounting. This in turn implies that it is possible to test the hypothesis of no mistakes (e.g., $\beta=1$ in the context of quasi-hyperbolic discounting) without considering anything other than choices. This statement seems inconsistent with the principle that it is impossible to falsify the principle of revealed preference with choice data alone. The explanation for this apparent inconsistency is that one

tests the hypothesis of no mistakes jointly with the assumptions of the structural model. Even if this joint hypothesis is rejected, there is some other structural model for which the hypothesis of no mistakes would not be rejected. When interpreting the results, one therefore necessarily relies on the non-choice evidence used to justify the assumed structure. Accordingly, the reliability and strength of this non-choice evidence limits the force of one's conclusions.

The observations in the preceding paragraph remain valid even if one uses data on non-standard types of choices, such as decisions made in advance of consequences, precommitments, and expenditures on self-control. For any given structural decision-making model, this type of evidence may prove extremely useful from the perspective of estimating parameters precisely and convincingly, and it may allow one to reject the hypothesis of no mistakes for a much broader class of preferences (e.g., any preference for which the decision-maker would exhibit time-consistent behavior). However, stepping outside of the assumed structure, there will always be other formulations of preferences that can explain the choice data without assuming a divergence between preferences and decisions. Of course, any such formulation will necessarily diverge from the standard model (as in Gul and Pesendorfer [2001]), and, in any given case, rationalization of the data may require strange assumptions about preferences.

It is worth emphasizing that the estimation of separate positive and normative models does not require us to abandon the principle of revealed preference completely. Instead, one implicitly invokes a principle of *selectively* revealed preference. Depending on the structural model, identifiable decisions (e.g., in the context of quasi-hyperbolic discounting, choices well in advance of consequences) may, by assumption, reveal preferences with certainty, or there may be uncertainty as to whether a given decision conforms to preferences (as in models with stochastic mistakes). In the latter case, one can model this uncertainty explicitly, proceeding, for example, as in the literature on switching regimes.

Identification of preferences through both choice and non-choice data. Another largely unexplored possibility would involve the use of both choice and non-choice data in structural estimation. Data of potential interest could include self-reported information about preferences and/or well-being, as well as measures of physical states such as arousal and stress.

This additional data could facilitate more precise and reliable estimation of key structural parameters. One might, for example, use self-reported data on preferences along with choice data to estimate the parameters of a normative model. In principle, the normative model could even include parameters that do not appear in the positive model. Likewise, non-choice data might prove useful in identifying circumstances in which choices reliably reflect preferences, and those in which they do not. If, for example, there is reason to believe that people are more prone to

make mistakes when they are under stress, data on cortisol levels might help to identify choices that more reliably reveal preferences.

The use of non-choice data raises at least two concerns. First, one can interpret this data through the lens of structural modeling only if one is willing to make additional assumptions, for example about how the non-choice data relate to decision-making processes. Advocates of the revealed preference approach view these assumptions with considerable suspicion (Gul and Psendorfer [2001,2004a,b]). However, an emerging theme in Behavioral Economics is that it is possible to justify, defend, and test these assumptions through the careful use of data from psychology and neuroscience. Furthermore, in practice the revealed preference approach relies on assumptions that are not directly supported by choice data – e.g., structural estimation *always* entails untested restrictions on the form of preferences – and people have different opinions as to which of these assumptions are most "reasonable" in a given instance. To the extent we judge an assumption as reasonable based on evidence not involving choice, it behooves us to make the basis of our inference explicit, regardless of whether we follow the standard approach or a behavioral alternative. One cannot claim an advantage for the standard approach simply by sweeping the implicit reliance on non-choice evidence under the rug, or by theorizing about an idealized procedure that is impossible to follow in practice (see Bernheim and Rangel [2005a] and Koszegi [2002] for elaborations of this point).

Second, economists generally view non-choice data as significantly less reliable and considerably more ambiguous than information on choices. In part, this view is justified by evidence indicating that certain types of self-reported data are unreliable (Diamond and Hausman [1994], Schwarz and Strack [1999]). In our view, this deficiency is exaggerated, particularly with regard to evidence concerning limitations and malfunctions of specific brain processes involving forecasting, memory, and attention (as discussed above). There is every reason to believe that the quality of this and other non-choice evidence, as well as our ability to interpret it, will improve with time. Furthermore, given the potential value of non-choice data, concerns about the quality of this information should motivate the development of better procedures for acquiring and interpreting it, rather than a policy of ignoring it on "conceptual" grounds.

We conclude this section by acknowledging two concerns. First, the feasibility and value of the empirical approach to measuring welfare discussed in this section has yet to be established through a series of persuasive applications. Only a few studies (discussed below) have made a start in this direction. There are many unresolved issues, e.g., concerning how to elicit and use data on self-reported preferences. Nevertheless, at a conceptual level, it does appear that one can

meaningfully conduct empirical welfare analysis allowing for some types of divergences between preferences and choices.

Second, there are significant political dangers associated with the research agenda described in this section. As we mentioned in Section 1, revealed preference is an attractive political principle because it prevents critics of any particular choice (e.g., concerning literature, sexual orientation, or religion) from condemning it on the grounds that it is contrary to a "natural" welfare criterion reflecting the individual's "true" interests. While we do not condone casual departures from this principle, we do think it is possible to insist on a high standard of proof, based in scientific evidence. In classifying certain behavioral patterns, such as psychoses, eating disorders, and addiction, as mental illnesses, the medical profession has grappled with essentially the same issues. While there have certainly been some dubious decisions (e.g., the classification, until relatively recently, of homosexuality as a psychiatric disorder), the process has, on the whole, reflected the balanced application of sound scientific principles.

3. Saving

For more that fifty years, the framework of intertemporal utility maximization has dominated economists' thinking about personal saving. This framework traces its roots to Irving Fisher (1930), and lies at the heart of the Life Cycle Hypothesis articulated by Modigliani and Brumberg (1954). In recent years it has become controversial, and an increasing number of economists have expressed doubts concerning its general validity. Many have turned to new approaches.

In this section, we survey some of the pertinent empirical evidence motivating the growing interest in alternatives to the standard model, describe some leading behavioral models, and explore some of their key policy implications. Our objective is to cover central themes. Given the size and rapid growth of this literature, we make no attempt to be comprehensive. Also, in describing competing models of saving, we focus on basic formulations, and ignore complications arising from liquidity constraints, intertemporal complementarities, and uncertainty about length of life and market parameters.

3.A. The policy issues

The last few decades have witnessed sharp declines in rates of saving for many developed countries. For example, according to statistics from the National Income and Product Accounts

(NIPA), the rate of net national saving for the U.S. dropped from 8.3 percent of net national product in 1980 to 1.8 percent in 2003. Low rates of saving have created widespread concern over investment, growth, the balance of payments, and the financial security of individual households. As a result, policymakers worldwide have become increasingly interested in developing strategies for stimulating thrift.

Public policies affecting private saving are highly contentious. In the U.S., policy makers are currently debating a variety of critical questions: Should the US partially replace its traditional social security system with individual savings accounts? If so, how should we structure the new system? Should the government impose more stringent regulations on defined contribution pension plans, which appear to be replacing defined benefit plans at a steady rate? Should we create or expand tax-deferred savings accounts for special needs, such as medical care and education? Or should we consider more fundamental tax reform that would reduce or eliminate the tax burden on capital income across the board?

To answer these and other critical questions, public economists require a theory of personal financial decision making that can explain observed behavior and generate credible out-of-sample predictions. It must also provide clear answers to normative questions, such as whether people save enough for retirement, and whether they invest their savings wisely.

3.B. The neoclassical perspective on saving

We begin by reviewing a simple version of the standard model. An individual lives for T+1 periods. In each period $t = 0,\ldots,T$, he consumes c_t units of an aggregate consumption good. His preferences are defined over consumption bundles of the form $c = (c_0,\ldots,c_T)$. We assume that it is possible to represent these preferences with a separable utility function of the form

$$U(c_1,...,c_T) = \sum_{t=0}^{T} \delta^t u(c_t),$$

where δ is a constant rate of time preference. The individual selects a consumption bundle from some feasible set, which reflects the distribution of earnings over time, interest rates, liquidity constraints, and the like. In practice, he chooses each element of c sequentially, rather than selecting the entire bundle at time 0. However, as time passes, he continues to apply the same lifetime preferences. This means that, as of time s, he evaluates continuation bundles, $(c_s,\ldots,c_T)$, according to the utility function

$$U_s(c_s,...,c_T) = A\sum_{t=s}^{T}\delta^{t-s}u(c_t) + B,$$

where $A = \delta^s$ and $B = \sum_{t=0}^{s-1}\delta^t u(c_t)$.

When writing down this model, economists usually follow the convention of renormalizing utility so that $A = 1$ and $B = 0$ in every period. This normalization obscures the fact that the individual has the same lifetime preferences at every moment in time. Since lifetime preferences are fixed, the appropriate welfare standard is unambiguous. Behavior is dynamically consistent in the sense that, fixing $(c_0,\ldots,c_{t-1})$, he would choose the same continuation bundle, $(c_t,\ldots,c_T)$, regardless of whether he made the decision in period t or some prior period. Accordingly, the individual behaves exactly as he would if he chose the entire bundle at time 0, which rules out any demand for precommitment.

The literature pertaining to the standard model is vast, and we make no attempt to review it here. However, in keeping with our objectives, it is important to summarize some of the key implications for public policy. The neoclassical approach assumes that people make appropriate decisions, provided they are well informed. If the government can provide relevant information more effectively and efficiently than private markets, educational policies are potentially beneficial. Assuming information is not an issue, there is no role for government in the absence of pre-existing distortions. It may be appropriate for the government to tax or subsidize capital income as part of a second-best policy in the presence of revenue requirements, to ensure an adequate level of competition in financial markets, to minimize fraud, and to alleviate adverse selection problems. However, under the standard view, there is nothing wrong with the choices people make, given the constraints they face. Reasons for government intervention involve market failures, not individual decision-making failures.

In practice, policy makers worry that people are not saving enough for their own security and future well-being. This is part of the motivation for proposals involving subsidized saving and/or mandatory accumulation. The standard model does not, however, recognize the legitimacy of this concern (except insofar as it results from a market failure). Under this view, saying that someone saves "too little" is comparable to asserting that he or she doesn't listen to enough classical music – thrift is simply a matter of taste (Lazear [1994]). In contrast, if households potentially make systematic mistakes, the adequacy of saving becomes a well-posed and important empirical issue.

In the ensuing sections, we review some of the evidence that calls the legitimacy of the standard approach into question, and we explore the implications of several emerging alternatives.

3.C. Some problematic observations

In some respects, saving behavior conforms reasonably well to the predictions of the Life-Cycle Hypothesis. For example, most people tend to accumulate wealth, broadly defined to include things like pension and social security entitlements, over the course of their working lives, and use either some or all of it to finance consumption after retirement. Yet there are also sound reasons to question the general applicability of this model and to examine alternatives. Here we list a number of problematic patterns identified in the literature. While it may be possible to account for some of these within the context of the Life-Cycle framework, collectively they pose a serious challenge to this approach.

1. Changes in consumption near retirement. The standard framework implies that people should smooth consumption, avoiding sudden and predictable changes in living standard. Yet a variety of studies have found that consumption declines sharply at retirement, when households experience a predictable decline in disposable income (Hammermesh [1984], Mariger [1987], Hausman and Paquette [1987], Robb and Burbridge [1989], Banks et. al. [1998], Bernheim et. a. [2001]). The decline in consumption is strongly correlated with accumulated wealth; those who accumulate less experience larger declines (Bernheim et. al [2001]).

One can try to account for this pattern within the standard model in several ways. First, retirement may be associated with a decline in work-related expenses and/or consumption goods that are substitutes for leisure. If these effects are anticipated, and if their magnitudes vary across the population, then people who plan for larger spending cuts after retirement will intentionally accumulate less wealth. Yet the evidence does not support this interpretation, as the effect is equally strong for categories of spending that would appear complementary to leisure and unrelated to work (Bernheim et. al. [2001]). Second, for those who stop working earlier than expected (e.g., due to disability), retirement reflects "bad news" to which consumption must adjust. Moreover, these same individuals find themselves with less-than-average wealth at retirement. However, even when the effects of unexpected retirement are removed through statistical procedures, one still observes both a decline in consumption at retirement, and a strong correlation between the size of this effect and accumulated wealth.

Notably, the sharp drop in consumption at retirement is also larger for households with lower rates of income replacement from social security and pension plans (Bernheim et. a.

[2001]). Once again, this pattern is observed even when the effects of unexpected retirement are removed. Since income replacement rates are easily anticipated, and since this variable is not likely to be strongly correlated with work-related expenses or a preference for leisure substitutes, standard theory is hard-pressed to account for the evidence.

This evidence would appear to indicate that people reduce consumption at retirement because they are surprised, either by the decline in their disposable income or by the inadequacy of their accumulated wealth. Yet other evidence suggests that the decline in consumption at retirement is anticipated (Hurd and Rohwedder [2003]). The explanation for this apparent puzzle remains an open question.

2. Self-reported mistakes. Several studies document large gaps between self-reported behavior and self-reported plans and/or preferences. A large fraction of the population reports saving too little – that is, significantly less than planned, or less than appropriate – for retirement (Bernheim [1995], Farkas and Johnson [1997], Choi et. al. [2004]). The reported gap is quite large, and few people report saving too much. Of those who express an intention to increase their saving, only a small fraction follow through (Choi et. al. [2004]). Taking these self-reports literally, one would conclude that pro-saving policies are potentially welfare-improving.

Skeptics counter that people are inclined to report "ideal" or "virtuous" behavior in answer to questions about plans or preferences; they might well also report that they watch too much television. This is a serious concern. However, the finding appears to be robust across samples, contexts, and phrasing of the pertinent questions. While the evidence is imperfect, in our view it should not be dismissed.

Others minimize the significance of the self-reported savings gap on the grounds that carefully calibrated life-cycle models can replicate data on wealth accumulation (see, e.g., Scholz et. al. [2004]). We find this line of argument unconvincing. At most, it supports an "as-if" interpretation of the life-cycle model. This does not rule out the possibility that people actually do make mistakes. Within the standard framework, one can rationalize a systematic tendency to consume too much as impatience – that is, a low value of δ. However, if overconsumption is indeed a mistake, then the true value of δ is higher than the as-if value, and this rationalization leads to an inappropriate welfare criterion. In addition, the models used to "explain" the level and distribution of wealth have other counterfactual implications (e.g., they produce no decline in consumption at retirement).

3. Limited planning skills. Most people are poorly equipped to engage in life cycle planning without assistance. Collectively, existing studies paint a rather bleak picture of economic and financial literacy (see, e.g., Walstad and Soper [1988], Walstad and :Larsen [1992],

O'Neill [1993], Consumer Federation of America and the American Express Company [1991], and Bernheim [1998]). For example, only 20 percent of adults can determine correct change using prices from a menu, and many have trouble determining whether a mortgage rate of 8.6 percent is better or worse than 8 ¾ percent. People tend to underestimate the power of compound interest, and many poorly understand common financial instruments.

In principle, financially illiterate individuals could seek guidance from experts. In practice, somewhere in the neighborhood of 60 percent of virtually every population subgroup relies primarily on parents, relatives, friends, and personal judgment. People with less education are actually *more* likely to rely on their own judgment. Only a minority consults financial professionals or print media (Bernheim [1998]). Moreover, in some cases financial professionals rely on simple rules of thumb (Doyle and Johnson [1991]), and even their relatively sophisticated tools conflict in some ways with sound life-cycle planning principles (Bernheim et. al. [2002]).

Financial literacy is strongly related to behavior. Those who are less financially literate also tend to save less (Bernheim [1998]). Moreover, measures designed to address financial illiteracy appear to have significant effects on choices. Policies mandating financial education for high school students result in higher asset accumulation once exposed students reach adulthood (Bernheim, Garrett, and Maki [2001]). Likewise, financial education in the workplace increases participation in employee-directed pension plans and stimulates saving (see Bernheim and Garrett [2003], Bayer, Bernheim, and Scholz [1996], and Duflo and Saez [2003]).

4. Failure to formulate sophisticated plans. Under an "as-if" interpretation, the standard model implies nothing about the process by which an individual arrives at consumption and saving decisions. Yet it is difficult to see how someone would formulate coherent life-cycle choices without extensive and deliberate planning. In practice, many people report spending little if any effort formulating long-range financial plans; moreover, those who fail to plan tend to save less (see Bernheim [1994], Lusardi [2000, 2003], Ameriks, Caplin, and Leahy [2003]).

When they exist, financial plans tend to be relatively unsophisticated. Many people establish saving targets, and in most cases think of these targets as percentages of income. However, the targets appear to reflect rough rules of thumb – in the vast majority of cases, they are integer multiples of 5 percent, and they vary neither with stated expectations about earnings growth nor with age (Bernheim [1994]).

In addition, important financial decisions often appear to turn on arguably irrelevant considerations. People are significantly more likely to make tax-deductible IRA contributions if

they owe the IRS money at the end of the tax year (Feenberg and Skinner [1989]).[4] There is a striking tendency for household to make an IRA contribution equal to the single-person limit, even when they are eligible to contribute more (Feenberg and Skinner [1989], Engen, Gale, and Scholz [1994]). And IRA participation rates rose sharply when the system was expanded in 1982, *even among groups that had been eligible prior to the expansion*, and fell sharply once the system was scaled back in 1986, *even among groups that remained eligible* (Long [1990], Venti and Wise [1992]).

5. The importance of default options. We use the term "default option" to signify the outcome resulting from inaction. For a neoclassical consumer, choices depend only on preferences and constraints. Consequently, in the absence of significant transaction costs, default options should be inconsequential. Yet in the context of decisions concerning saving and investment, they appear to matter a great deal.

With respect to 401(k) plans, there is considerable evidence that default options affect participation rates, contribution rates, and portfolios (Madrian and Shea [2001], Choi et. al. [2004a]). Also, automatic cash distributions for terminated employees with small balances reduce retirement account balances, even though these employees are free to roll their funds into an IRA (Choi et. al. [2004a]).[5] Effects of defaults on portfolio allocation have also been documented in the context of the recent privatization of social security in Sweden. The dissemination of information about investment alternatives appears to counter this effect (Cronqvist and Thaler [2004]).

In the standard framework, defaults can matter if other choices are associated with significant transaction costs. Yet in the contexts described above, transactions costs are presumably quite low. Alternatively, the effect of a default option may be related to the costs of decision making. In pressing this explanation, one must explain why these costs favor the default option over other alternatives (e.g., the simplest or most transparent choices). One possibility is that people believe the default conveys information about the wisdom of a particular choice. This may be a plausible assumption in the context of portfolio allocation within 401(k) plans, where the employer has a fiduciary responsibility to its employees in its role as plan sponsor. In any case, even if default options are viewed as informative, their strong effects tell us that people regularly make significant decisions concerning saving on the basis of precious little information.

---

[4] Gravelle [1991] attributes this to spurious correlations with income, tax filing status, and/or asset holdings, but the pattern is apparent even when Feenberg and Skinner include plausible controls for these factors.

[5] Choi et. al. [2004a] also contains a discussion of the "optimal defaults".

6. Inefficient choices. In the standard framework, consumers always choose alternatives on the efficient frontiers of their constraint sets. When evaluating evidence pertaining to this implication, it would be unfair to interpret it too literally. In some instances (e.g., failure to engage in sophisticated tax arbitrage), squeezing out the last dime involves complex arrangements and potentially high transaction costs, so the appearance of inefficiency may be illusory. However, in some cases, people select alternatives far from the efficient frontiers of their choice sets in settings where superior alternatives are clearly available. Examples include failures to take advantage of low interest loans available through life insurance policies (Warshawsky [1987]), naïve diversification strategies (Bernartzi and Thaler [2001]), the tendency to invest 401(k) balances heavily in the stock of one's employer (Holden, Van Der Hei and Quick [2000] and Bernartzi [2001]), the proclivity to maintain substantial balances on high-interest credit cards (Laibson et. al. [2003], Laibson et. al. [2004], Gross and Souleles [2002]), and the inclination to delay IRA contributions until the end of the tax year (Summers [1986]).

3.D. Insights from psychology

A number of the empirical puzzles described in the previous section may be related to problems involving the exercise of self-control. There is a sizable and rapidly growing literature in psychology and neuroscience concerning the properties, development, and limitations of self-control processes. In this section we provide a brief introduction to this literature by summarizing some of the evidence most relevant for savings. See Frederick, Loewenstein and O'Dohonue [2002] and Loewenstein, Read, and Baumister [2003] for more comprehensive reviews of the literature.

Evidence of dynamically inconsistent choice. Saving reflects a decision to accept a lower level of consumption in one period in exchange for a higher level of consumption in another. The standard model assumes that the individual evaluates a tradeoff involving consumption at two future fixed points in time, say s and t (with $s < t$), precisely the same way at every moment r. Yet a large body of evidence finds that this evaluation in fact depends on the proximity of r to s. In particular, when s is sufficiently proximate, people tend to favor consumption in the closer period s.

The direct evidence for this proposition is experimental. The typical experiment involves two treatments. In the first, subjects are offered a small prize in s days, or a large prize in t days. In the second, they are offered the same small prize in $s + d$ days, or the same large prize in $t + d$ days, for some $d > 0$ (where we interpret d as "delay"). When s = 0 (that is, the subject decides between an immediate reward and a delayed one in the first treatment), a significantly larger

fraction of subjects choose the small prize in the first treatment than in the second (see, e.g.., Ainslie and Haendel [1983], or, for a recent review of the evidence, Frederick, Loewenstein, and O'Donoghue [2002]). For relatively small values of *s* (on the order of seven days), this differential disappears (Harrison, Coller, and Rutstrom [2002]).

The simple experiment described in the previous paragraph potentially suffers from a variety of confounds. An immediate reward is usually distinguished by more than just its immediacy. Arguably, it is less risky (that is, less likely to be forgotten by the subject or neglected by the experimenter), and it involves lower transaction costs. However, the discrepancy between the two treatments persists even when reasonable steps are taken to eliminate these confounds. Another concern is that, with state-contingent utility, evaluations of tradeoffs may depend on "moods." For an immediate reward, mood is known, while for a future reward it is not. Under appropriate (if somewhat special) assumptions, this can account for the observed pattern (Fernandez-Villaverde and Mukherji [2002]).

Notably, similar results are obtained regardless of whether the reward consists of money or a consumption good. This is surprising in that, for a wide range of standard and non-standard behavioral theories, the best choice with monetary rewards involves the maximization of present discounted value (at least in the absence of binding liquidity constraints), which means it should not vary with delay, *d*.

Pre-commitment. People who understand that their behavior is dynamically inconsistent might want to exercise self-control through the use of pre-commitment devices. There is evidence that this occurs in practice. For example, Ariely and Wertenbroch [2002] study a field experiment in which students are allowed to self-impose deadlines on assignments. They find that many subjects choose these constraints. Wertenbroch [1998] discusses suggestive evidence that people attempt to control their consumption of "tempting" foods by purchasing small packages, even when the unit price is lower for larger packages.

The role of cues and cognitive processes in self-control. In an influential study, Shiv and Fedorihin [1999] show that cognitive load can affect self-control. Subjects are given a number to memorize, and are asked to report it in another room. In some cases the number has two digits, and in others it has seven. Before reporting the number, they are asked to choose between two deserts, chocolate cake and fruit salad, which are physically present. Individuals in the seven-digit treatment are roughly 50% more likely to choose the chocolate cake. This suggests that self-control requires cognitive effort, and that this becomes more difficult when cognition is engaged in other tasks.

Shiv and Fedorihin [1999] also consider a variation of this experiment in which the deserts are not physically present; instead, subjects are shown pictures. The differential in choices between the two treatments disappears. This suggests that cues can impair self-control. To account for this effect, psychologists hypothesize that self-control is difficult when the individuals enter strong "visceral states," and that the real items are more likely than pictures to trigger such states.

These findings are consistent with the work of Mischel and co-authors, which shows that self-control is affected by the deployment of attention and the presence of cues (see Mischel [1974], Mischel and Moor [1973], Mischel, Shoda, and Rodriguez [1992] and Metcalfe and Mischel [1999]). In a typical experiment, a subject (often a child) is placed in a room and is offered a choice between an inferior and a superior prize (one or two pieces of candy). Subjects can obtain the inferior prize at any time by calling the experimenter, but must wait until he returns to obtain the superior prize. In practice, the child's ability to wait depends crucially on whether the inferior prize is visible. Merely covering the object significantly enhances self-control.

More generally, in Mischel's experiments, the deployment of attention emerges as a key determinant of self-control. Any stimulus that focuses attention on the "tempting" features of the inferior prize increases the likelihood that the children will select it. Children are significantly more likely to wait if they are advised to distract themselves by thinking about something else, or if they are provided with a toy, even when children in a control group show no interest in the toy.

Discussion. The evidence suggests that exercising self-control is sometimes difficult. The amount of effort devoted to imposing self-control appears to depend on a variety of environmental and contextual factors that are arguably unrelated to true preferences. Accordingly, lapses in self-control are potentially associated with divergences between choices and true preferences (i.e., mistakes). Moreover, one expects such lapses to arise probabilistically, as the result of chance encounters with cues and stimuli outside the individual's control.

The models of decision making described in the next two sections attempt to capture these ideas in different ways. They make different assumptions about the nature of the processes responsible for the mistakes associated with self-control lapses, and they employ different reduced-form representations of these processes.

3.E. Models of saving with quasi-hyperbolic discounting

Building on previous work by Strotz [1956], Phelps and Pollack [1968], and Akerlof [1991], Laibson [1997] proposes a model of saving intended to capture some of the self-control problems described in Sections 3.C and 3.D. This framework is widely known as "quasi-

hyperbolic" or "(β,δ)" discounting.[6] From a positive perspective, individuals behave as if they optimize subject to lifetime preferences that change with time. In particular, in each period *t*, the decision maker acts as if he picks the feasible consumption path that maximizes a utility function of the form

$$u(c_t) + \beta[\sum_{k=t+1}^{T} \delta^{k-t} u(c_k)].$$

This formulation differs from the standard model in only one respect: it includes an additional discount factor, $\beta > 0$, that is applied to the utility associated with all future consumption. The parameter β is meant to represent the degree of *present bias*, or *myopia*. The standard model corresponds to the special case where β=1. With $\beta < 1$, the present is given special status relative to all other time periods, and this creates a powerful tendency to consume immediately.

As long as $\beta \neq 1$, this model gives rise to dynamically inconsistent behavior. With $\beta < 1$, the individual always wishes to consume more in the current period than he would have chosen for himself at any point in the past. This complicates positive analysis. One can no longer characterize the individual's behavior by solving a single optimization problem. Instead, the model gives rise to a game played between "multiple selves." The literature solves this game under three different assumptions about the accuracy of the decision maker's expectations concerning his own future behavior.

A naive individual acts as if his future selves will be willing to follow through on his current plans. In this case, one determines behavior by solving a sequence of optimization problems. In each period, the naïve self divides his resources between current consumption and saving, anticipating that he will use his wealth to finance his desired consumption path for the rest of his life. He never actually follows this plan because, in the next period, he again attaches disproportionate weight to the present. The naïve individual does not understand his self-control problem, and makes no attempt to manage it.

A sophisticated decision maker perfectly anticipates his future actions. In particular, he knows that, given the opportunity in any future period, he will consume a larger fraction of his resources than he would like. Under this assumption, one determines behavior by solving for the sub-game perfect equilibria of the dynamic game played between multiple selves. Frequently, this setting gives rise to multiple equilibria, which means behavior is indeterminate unless one applies a selection criterion or refinement (Laibson [1994], Krussel and Smith [2003], and Bernheim, Ray, and Yeltekin [1999]). In contrast to naïve decision makers, a sophisticated

[6] See O'Donohue and Rabin [1999a,b] for other early influential variations of the (β,δ) model.

decision maker perfectly understands his self-control problem, and may attempt to manage anticipated lapses of self-control by limiting future choices.

Finally, a partially sophisticated decision maker understands that he will have a self-control problem in the future, but underestimates its magnitude. O'Donoghue and Rabin [1999b,2001] parameterize the degree of sophistication to create a continuum between the two extreme cases of complete naivete and perfect sophistication. See their papers for details, as well as for further discussion of the relationships between these assumptions.

There has been much confusion in the literature concerning interpretations of the (β,δ)-model. This confusion reflects the fact that the positive model described above is consistent with at least two distinct approaches to the formulation of a normative model. One approach follows the agenda outlined in section 2.C: think of person A at time *t* as the "child" of person A at time *t-1*, and then apply standard multi-person welfare principles. The second approach follows the agenda outlined in section 2.D: assume the individual has stable lifetime preferences, and interpret the reduced-form parameter β as measuring the tendency to make present-biased mistakes. With few exceptions, the leading advocates of the (β,δ)-model endorse the second approach.[7] Typically, they assume that true preferences correspond to a standard intertemporal utility function with exponential discounting at the rate δ ("long-run" preferences).[8] Yet much of the profession continues to think of the (β,δ)-model literally as one with "multiple selves," which is in keeping with the first approach, but not the second.

Several papers have estimated (or calibrated) (β,δ) models using data on consumption and saving. In principle, this permits one to test the hypothesis that $\beta=1$. Under the second approach to normative analysis described in the preceding paragraph, it also allows one to recover true preferences, and to conduct welfare analysis.

Angeletos et. al. (2001) simulate a 90 period life-cycle model with uncertain labor income, probabilistic death, constant discount factors, additively separable preferences, and three types of assets: riskless bonds, credit card borrowing, and an illiquid asset resembling housing wealth. They calibrate the model to match the median level of wealth near retirement assuming $\beta = 1$, and again assuming $\beta = 0.7$. Then they compare the model's ability to track data from the Panel Study of Income Dynamics (PSID) under these two different assumptions. Both versions generate similar consumption patterns, except that borrowing is higher earlier in life and

[7] This statement is based in large part on personal conversations. Much of the literature is not explicit on this point.

[8] As discussed in Section 2.C, one can justify the same welfare criterion under the first approach.

consumption is higher later in life with quasi-hyperbolic discounting. However, with $\beta = 0.7$, the model performs substantially better in tracking credit card balances, the share of wealth held in liquid form, the marginal propensity to consume out of anticipated income, and the discontinuity in consumption at retirement.

Laibson et. al. [2004] develop and estimate a similar model with stochastic labor income, liquidity constraints, child and adult dependents, liquid and illiquid assets, and revolving credit. They use the Method of Simulated Moments to estimate many of the parameters of the model based on data from the Survey of Consumer Finances. They formally reject the standard exponential model in favor of quasi-hyperbolic discounting. According to their estimates, the short-run annualized discount rate is 40%, while the long-run annualized discount rate is only 4%. Their rejection of exponential discounting is driven by the observation that high levels of credit card borrowing coexist with significant wealth accumulation. Paserman (2002) uses labor market data on unemployment durations and market wages to estimate a related model. He finds a long-run discount rate of 0.1% and a short-term discount rate of 10-60%. Fang and Silverman (2002) conduct a similar exercise using welfare participation data.

These studies exemplify the approach to empirical Behavioral Public Economics described in section 2.D. They demonstrate the feasibility of this approach, and provide important evidence in support of a behavioral approach to savings policy. However, much additional empirical work is required to establish the stability, robustness, and scope of these findings.

It is important to emphasize that, while this collection of empirical papers provides evidence against the standard model, they do not allow one to conclude that the (β,δ) model outperforms other behavioral alternatives, such as those discussed in the ensuing sections. The patterns in the data that produce estimates of β less than unity could result from other processes that generate excessive consumption. To our knowledge, no one has yet undertaken empirical comparisons of alternative behavioral models.

The policy implications of the (β,δ)-model are dramatically different from those of the standard model. Since many individuals choose sub-optimally low levels of saving, there may be welfare improving policy interventions *even in the absence of capital market failures*. First, mandatory savings programs may be welfare-enhancing, provided they are large enough to crowd out private savings (in the form of liquid assets) at some point during the life cycle (Imrohoroglu et. al. [2003]). See Feldstein (1985) for a characterization the optimal level of social security benefits in an overlapping generations economy with two-period lifetimes and heterogenous self-

control problems, and Diamond and Koszegi [2003] for an analysis of social security with quasi-hyperbolic discounting and endogenous retirement.[9] Third, as long as the population includes some individuals with self-control problems, and assuming the social welfare function is continuous and concave, a small subsidy for saving financed with lump-sum taxes is welfare improving. Intuitively, since individuals with self-control problems save too little, the subsidy produces a first-order improvement in their well-being, and has only a second-order effect on the welfare of those without self-control problems. For a discussion of optimal taxation in the (β,δ)-model, see O'Donoghue and Rabin [2005] and Krusell, Kuruscu, and Smith [2000,2002]. Finally, introducing restrictions on the availability of credit, for example by regulating the distribution of revolving credit-lines and mandating credit ceilings, can significantly enhance the well-being of those with self-control problems.

3.F. Models of savings with cue-triggered mistakes

Bernheim and Rangel [2005b] propose an alternative model of savings in which individuals make stochastic mistakes. As in the standard model, true preferences correspond to an additively separable function with exponential discounting. The individual makes decisions in two distinct modes. With probability p_t, decision processes function properly, and he optimizes as in the standard model. With probability $1- p_t$, decision processes are in faulty (implicitly because an environmental cue triggers a lapse of self-control), and he consumers excessively. He can influence the probability of encountering cues that trigger the faulty decision mode through choices of activities (for example, whether to shop at expensive stores).

In the functional mode, the decision-maker is sophisticated about his self-control problem: he selects the optimal level of current consumption recognizing the probabilities and consequences of entering the faulty mode in the future, as well as the manner in which his actions affect the distribution of future decision modes. In the faulty mode, he "binges." This response is mechanical, reflecting simple impulses. In the simplest versions of the model, the size of the binge is proportional either to intended consumption (e.g., because he has chosen to shop in an expensive store), or to remaining lifetime resources (where the factor of proportionality is sufficiently large to ensure that the binge exceeds intended consumption). In either case, the size of the binge is constrained by his available liquid resources.

The model has two straightforward implications. First, pre-commitment technologies are valuable because they can reduce size of a mistake when the faulty mode is triggered. Second, the

[9] Feldstein does not use the (β,δ) language, but his model is a special case of this framework.

consumer can actively manage his self-control problem, for example by choosing activities that reduce the likelihood of encountering cues that trigger binges. If the size of the binge is related to intended consumption, he can also reduce the size of mistakes, when they occur, by planning to consume less (e.g., lapses are less costly if he shops at less expensive stores).

Other implications of the model are less immediate. While an increase in the probability or size of a binge always reduces welfare, it can either increase or decrease the level of saving (depending on parameter values). Additional saving becomes more attractive because it allows the individual to self-insure against future mistakes. However, it also becomes less attractive because it leads to greater waste. The net effect on savings depends on the balance of these two forces.

The model also predicts the existence of low-asset traps. For an individual with few assets, the size of a binge is constrained by liquid resources. If he saves an additional dollar and then experiences a binge, the entire dollar is wasted. For an individual with substantial wealth, the size of a binge is ordinarily not constrained by liquid resources. If he saves an additional dollar and then experiences a binge, only a fraction of the dollar is wasted. Consequently, saving is relatively less attractive when wealth is low.

With respect to durable consumption goods, the implications of this model potentially differ from those of the (β,δ)-framework. The (β,δ)-model envisions present-bias with respect to consumption flows. Consequently, it cannot explain excessive consumption of durable goods with long lives, for which the bulk of consumption occurs in the future. In contrast, since an individual may act impulsively with respect to both present and future consumption, a model with stochastic cue-conditioned decision modes can easily generate excessive consumption of durable goods. Accordingly, this model potentially justifies cooling-off periods for automobile purchases, whereas the (β,δ)-model does not.

Many of the policy implications of this model parallel those (β,δ)-framework. Even in the absence of capital market imperfections, government intervention is potentially welfare-improving. The introduction of mandatory savings can enhance the well-being of those with self-control problems, but only if the program is large enough to crowd out all liquid assets at some point during the life-cycle, in some state of nature. Regulations that restrict the availability of credit are also potentially beneficial.

There are, however, important differences between the two models. Perhaps most notably, whereas optimal policy in the (β,δ)-model entails subsidized savings, in this model either taxation or subsidization of saving may be optimal. To understand why, note that there are two

key differences between the models. First, in the (β,δ)-model consumers *always* make present-biased mistakes, while in this model mistakes are stochastic. This means that social insurance considerations come into play. To partially insure the consumer against bad realizations, the government should give him money when random events reduce his wealth, and take money away when random events increase his wealth. In this context, the random event that potentially reduces his wealth is a cue-triggered binge. A capital income tax (coupled with a lump-sum subsidy) supplements the individual's wealth when he experiences a binge (because his saving is low), and reduces his wealth when he does not binge (because his saving is high). Second, in the (β,δ)-model, the decision maker responds to future economic incentives even while making mistakes, whereas this model assumes that errors result from a mechanical and largely inflexible impulses. Accordingly, taxation directly reduces the magnitude decision errors in the (β,δ) framework, but has a limited effect on binges in this model. [10]

In models with cue-triggered binges, there is also a natural role for cognitive policies such as the regulation of advertising and marketing. If advertising increases the likelihood and size of mistakes by proliferating cues, restrictions on advertisements are potentially welfare-improving, particularly if their information content is small. However, for the reasons discussed above, the impact of such restrictions on the level of saving is ambiguous. One could incorporate the same forces in the (β,δ)-model by assuming that advertising reduces the value of β. In contrast to the current model, this would necessarily reduce saving (provided the consumer's horizon is finite).

One can also rationalize framing effects in this model by assuming that the probability of entering the faulty mode depends on cues embedded in the presentation of a decision problem. It may then be possible to design savings plans that increase thrift without providing new information or changing budget constraints, as claimed by Thaler and Shefrin [2004].

The model of savings described in this section is closely related to the process-malfunction theory of addiction discussed below in section 4.E. Since we advocate the use of reduced form models of decision making justified by evidence on underlying psychological and neural processes, we end this section with a disclaimer. In the context of addiction, the hypothesis that people make cue-triggered mistakes has a solid foundation in neuroscience. In the context of saving, the foundations are less solid. As emphasized in section 4.D, it is known that self-control plays a critical role in determining saving, and a significant body of evidence

[10] It is worth mentioning that the (β,δ) model also fails to explain an important general fact about present-bias – that the phenomenon persists even in experiments where participants are rewarded in dollars, rather than with rewards experienced at fixed points in time. Even a (β,δ) discounter should always maximize the present discounted value of resources.

suggests that cues influence the ability to impose self-control. However, it is difficult to draw a clear distinction between a lapse of self-control and, say, a temporary (and possibly cue-triggered) state of impatience. Our understanding of the neurobiology of self-control, and how it relates to intertemporal choice, is still preliminary.

3.G. Models of savings with non-standard preferences

Gul and Psendorfer [2004a,b] propose an alternative model to account for the role of self-control in determining saving. In contrast to the approaches discussed in the preceding sections, they adhere to the principle of revealed preference, thereby excluding the possibility that lapses of self-control involve mistakes. According to their model, the consumer acts as if he maximizes an intertemporal utility function of the following form:

$$U(c_1,...,c_T;B_1,...,B_T)=\sum_{t=0}^{T}\delta^t u(c_t,B_t),$$

where B_t denotes the budget set in period t. The inclusion of B_t as an argument of u differentiates this framework from the standard approach. The budget constraint enters preferences in a specific way:

$$u(c_t,B_t)=v(c_t)-[\max_{c\in B_t}\tau(c)-\tau(c_t)],$$

where $v(.)$, the flow of utility of consumption, and $\tau(.)$, the level of temptation associated with a given option, are increasing concave functions satisfying the usual properties. The second term (in brackets) reflects the unpleasant sensation of temptation experienced by the consumer when he fails to select the most tempting alternative in his budget set.

To understand how the model works, it is useful to consider a simple consumption-saving problem with two periods, no discounting, and zero interest. Let R denote the amount of resources available to the individual in period 1, and let $s=R-c_1$ denote the level of saving. The period 2 value function is given by

$$V_2(s)=v(s)-[\max_{c\in[0,s]}\tau(c)-\tau(s)]=v(s).$$

That is, since the individual spends all his resources in the second period, he does not experience unpleasant temptation. Using this expression, we can write lifetime utility as a function of first-period saving:

$$V_1(s)=v(R-s)-[\max_{t\in[0,R]}\tau(R-t)-\tau(R-s)]+V_2(s).$$

$$=v(R-s)-[\tau(R)-\tau(R-s)]+v(s)$$

In the absence of temptation, the individual would simply maximize $v(R-s)+v(s)$. At an interior solution, this requires $v'(R-s)=v'(s)$. The introduction of temptation increases the cost of savings by $\tau'(R-s)$, which causes saving to fall.

Several properties of the model are worth highlighting. First, the presence of temptation can decrease well-being even if does not affect behavior. In this sense, self-control is costly. Second, the individual is always (weakly) better off when a planner removes all discretion and forces him to consume the allocation that would be optimal in the absence of temptation. Third, the individual experiences temptation with respect to current choices, but not with respect to future choices. (He is not, for example, tempted to purchase a sports car delivered with some lag.) As a result, in the absence of uncertainty, an individual who has the ability to lock in choices one period in advance can achieve the first-best (except in the first period). Fourth, as in the standard model, choices are dynamically consistent.

Gul and Psendorfer's model can be interpreted as a reduced form representation of the process that generates the costs associated with temptation and the exercise of self-control. A closely related model, pioneered by Thaler and Shefrin [1981] and recently revisited by Fudenberg and Levine [2005], makes the sources of these costs more explicit. Preferences are given by an intertemporal utility function of the form

$$U(c_1,...,c_T;a_1,...,a_T)=\sum_{t=0}^{T}\delta^t u(c_t,a_t),$$

where a_t measures the intensity with which the individual deploys self-control in period t. The consumer chooses a_t at the outset of each period with the object of maximizing intertemporal utility; he then chooses c_t myopically, based on immediate benefits. The imposition of self-control is costly in the sense that $\partial u/\partial a<0$, but it leads to lower consumption.

As shown by Benabou and Pycia [2002], O'Donoghue and Loewenstein [2004], and Fudenberg and Levine [2005], this framework is equivalent over consumption-saving choices to Gul and Pesendorfer's theory of temptation. See also Loewenstein-O'Donoghue [2004] for an insightful discussion of the relationship between this class of models and the (β,δ)-framework.

Gul and Psendorfer [2004a,b] emphasize that their approach is conceptually consistent with the method of revealed preference. Supposedly, this eliminates the need for non-choice data, and prevents the policy analyst from imposing his or her own judgments when evaluating welfare. We disagree. Practical implementation of the revealed preference methodology requires the analyst to make assumptions about the data generating process (e.g. about functional forms, or similarities across individuals). There are always untested assumptions, which the analyst selects

based on other information, instinct, introspection, or fuzzy notions of "reasonableness." We believe it is fair to say that these assumptions are not chosen exclusively on the basis of choice data. Moreover, as all veterans of empirical policy debates are aware, the analyst's judgments about untested assumptions translate directly into judgments about welfare. There are also theoretical considerations, which we discuss at length in Bernheim and Rangel [2005b]. Assuming one restricts attention to data on choices over allocations and constraint sets, both the standard theory and Gul and Pesendorfer's model are observationally equivalent to other models with different welfare implications. Hence, the analyst's judgment, expressed through axioms and assumptions, is unavoidable.

What are the novel policy implications of the temptation model? First, mandatory savings programs can improve welfare *even if they do not increase savings*. This follows from the fact that any limit on consumption reduces temptation. In contrast to models with (β,δ) discounting and cue-triggered mistakes, a small program of mandatory saving can enhance welfare even if people still retain positive liquid assets in all time periods and states of nature. Second, unlike models with (β,δ) discounting and cue-triggered mistakes, there is no role for corrective taxation. See Krusell, Kuruscu, and Smith (2001) for further results and discussion.

3.H. Discussion

Economists have only recently begun to study saving using tools from behavioral economics. Even so, the models described in this section have already provided valuable insights. We conclude this section with a brief description of some important open questions.

The models described in this survey provide an explanation for some of the patterns described in sections 2.C and 2.D, including time inconsistency, self-reported mistakes, and some types of inefficient financial choices. However, it is not clear that they can adequately account for other patterns, such as the discontinuity of consumption near retirement, the role of default options, the failure to plan, and the use of rough rules of thumb. None provides a fully satisfactory explanation for the success of the Saving for Tomorrow Savings Plan™ designed by Thaler and Shefrin [2004], which relies on framing effects instead of changes in budget constraints. Nor do they incorporate limitations on financial skills. In focusing on self-control problems, they ignore issues associated with the complexity of financial decision-making.

Likewise, the theoretical work described in the previous sections has formalized only a few of the behavioral channels through which public policy could affect choices and welfare. It

is important to study other behavioral mechanisms with the same level of rigor. Interesting possibilities include the following.

1. The role of financial professionals. Many people rely on advice from financial professionals. One can therefore potentially learn about behavior by studying the methods used to generate this advice (see e.g. Bernheim et. al. [2002]). For example, the most common retirement planning technique involves setting some fixed target for retirement (usually derived from an arbitrary earnings replacement rate) and computing the annual inflation-adjusted contribution to savings sufficient to achieve this target (see Doyle and Johnson [1991]). This generates a negative interest elasticity of saving because higher rates of return make it easier to accumulate the resources required to reach the target.

2. Social influences. When saving incentives are in place, boundedly rational individuals may be more likely to learn that others regard the benefits of saving as important. For example, the availability of a 401(k) in an employment setting may stimulate conversations about contributions and investments, and thereby produce "peer group" influences involving both demonstration and competition (see, e.g., Duflo and Saez [2002, 2003]). The very existence of a pro-saving policy may indicate that "authorities" perceive the need for greater thrift, or endorse a particular level of saving (e.g., the contribution limit).

3. Keeping score. By segmenting retirement saving from other forms of saving, certain kinds of tax-favored accounts may make it easier to monitor progress towards long-term objectives. Information on total accumulated balances is usually provided automatically, or is readily available. This gives individuals a convenient yardstick for measuring the adequacy or inadequacy of their thrift. This may have the effect of making the costs of short-sightedness more explicit. It could also help people formulate goals and simple behavioral rules. According to Thaler and Shefrin (1981), "[s]imply keeping track seems to act as a tax on any behavior which the planner views as deviant."

4. Intrinsic motivation. Scitovsky [1976] has raised the possibility that some individuals may view saving as a virtuous activity in and of itself, without any explicit contemplation of future consequences (see also Katona [1975]). Pro-saving policies may promote this outlook by reinforcing the notion that, as something worthy of encouragement, saving is intrinsically rewarding and immediately gratifying.

5. Intrinsic gratification from tax avoidance. We have noted that people are more likely to contribute to IRAs if they owe money at the end of the tax year. This suggests that immediate tax avoidance is intrinsically gratifying. If so, "front-loaded" plans, wherein contributions are deductible and withdrawals are fully taxable, may be more effective in

stimulating saving than "back-loaded" plans, wherein contributions are not deductible and withdrawals of *principal* are not taxable.

6. Mental accounting. Shefrin and Thaler [1988] and Lowenstein and Prelec [1998] argue that people exercise self-control by separating resources into "mental accounts," each associated with a different objective. IRAs and 401(k)s may reinforce the discipline of mental accounting by earmarking certain resources for retirement, particularly in the presence of penalties for early withdrawal.

7. Education and promotion. The existence of tax-deferred savings accounts may stimulate promotional activities and advertisements by financial services firms. Policies that favor the development of employee-directed pensions (like 401(k)s) may encourage employers to provide retirement education. While advertising and education appear to affect financial decisions, the precise mechanisms are poorly understood.

These types of considerations potentially have important implications for critical policy questions, such as the choice between broad-based policies for promoting saving (e.g., consumption taxation) and more targeted strategies (e.g., IRAs). From a behavioral perspective, narrow measures can focus attention on a single issue (such as the adequacy of saving for retirement), expose individuals to information concerning the importance of saving, provide a natural context for the development and enforcement of private rules, and promote the growth of pro-saving institutions. Contribution limits may actually stimulate saving if they validate specific targets, provide natural focal points for the formation of private rules, or make it easier to monitor compliance with these rules.

4. Addiction

Although more than four million chemical compounds have been catalogued to date, only a few score are classified as addictive by clinical consensus (Gardner and David (1999)). These include alcohol, barbiturates, amphetamines, cocaine, caffeine and related methylxanthine stimulants, cannabis, hallucinogens, nicotine, opioids, dissociative anesthetics, and volatile solvents. There is also some debate as to whether other substances, such as fats and sugars, or activities, such as shopping, shoplifting, sex, television viewing, and internet use, are clinically addictive. These substances and activities pose challenges both for public policy, and for standard economic analysis.

This section reviews the distinctive behavioral patterns associated with the consumption of addictive substances, describes the neuroscientific foundations of addiction, summarizes several competing economic models, and reviews their policy implications.

4.A. The policy issues

The consumption of addictive substances raises important social issues affecting members of all socioeconomic strata, and citizens of virtually every nation. Readily available statistics for the United States illustrate the scope of the phenomenon.[11] Estimates for 1999 place total expenditures on tobacco products, alcoholic beverages, cocaine, heroin, marijuana, and methamphetamines at more than $150 billion. During a single month in 1999, more than 57 million individuals smoked at least one cigarette, more than 41 million engaged in binge drinking (involving five or more drinks on one occasion), and roughly 12 million used marijuana. In 1998, slightly more than 5 million Americans qualified as "hard-core" chronic drug users. Roughly 4.6 million persons in the workforce met the criterion for a diagnosis of drug dependence and 24.5 million had a history of clinical alcohol dependence. In 1998, additional social costs resulting from health care expenditures, loss of life, impaired productivity, motor vehicle accidents, crime, law enforcement, and welfare totaled $185 billion for alcohol and $143 billion for other addictive substances. Smoking killed roughly 418,000 people in 1990, alcohol accounted for 107,400 deaths in 1992, and drug use resulted in 19,277 deaths during 1998. Alcohol abuse contributed to 25 to 30 percent of violent crimes.

Even within jurisdictions, public policy toward various addictive substances is far from uniform, despite the commonalities suggested by their shared clinical classification. Policies range from laissez faire to taxation, subsidization (e.g. of rehabilitation programs), regulated dispensation, criminalization, product liability, and public health campaigns. Each alternative policy approach has passionate advocates and detractors.

Despite sharp disagreements about the ideal treatment of addictive substances, there is reasonably widespread agreement that most existing policies work poorly. The U.S. “War on Drugs” is, for example, often labeled a “failed policy.” Use of banned substances remains widespread, and the resulting health costs are high. Prohibitions on certain substances, like marijuana, lack credibility among younger Americans, who fail to see why alcohol is singled out as socially acceptable. While the incidence of criminal activity among drug addicts is relatively high, it is important to acknowledge that drug related-crime is, to a significant extent, a consequence of current policy, rather than a justification for it. Criminalization promotes black

[11] The statistics in this paragraph were obtained from the following sources: Office of National Drug Control Policy [2001a,b], U.S. Census Bureau [2001], National Institute on Drug Abuse [1998], National Institute on Alcohol Abuse and Alcoholism [2001], and Center for Disease Control [1993]. There is, of course, disagreement as to many of the reported figures.

markets, fosters organized crime, enriches criminals, and contributes to a culture of violence. As a result, more than 625,000 citizens were incarcerated for drug-related offenses curing 1999. These people were disproportionately poor, black, and among society's most economically vulnerable members.

While existing policies have serious drawbacks, alternatives are also potentially problematic. For example, the high incidence of alcohol abuse and smoking, along with the attendant social costs, at a minimum raise serious concerns about the potential consequences of across-the-board legalization. The apparent intractability of social problems related to addiction underscores the importance of creatively and openly rethinking policy strategies.

4.B. The neoclassical perspective on addiction

Prior to the 1990s, neurological theories of addiction were based on the "pleasure principle". It was widely believed that people start using drugs to achieve a pleasurable "high," and continue using them despite a deterioration of the high (a phenomenon known as "tolerance") to avoid unpleasant feelings associated with cravings and withdrawal. These hedonic properties are easily incorporated into standard models of consumer choice. Early work in this tradition includes papers by Stigler and Becker [1977], Iannacone [1986], and Becker and Murphy [1988]. The last of these is widely viewed as the definitive articulation of the neoclassical perspective on addictive behavior, also known as the theory of "rational addiction."

In Becker and Murphy's model, the individual's well-being depends on consumption of an addictive good, consumption of a non-addictive good, and a state variable summarizing past consumption of the addictive good. This addictive state rises with use of the substance and falls with abstinence. To model tolerance, one assumes that utility declines as the addictive state rises. To model the effects of cravings and the pain of withdrawal on the inclination to use a substance, one assumes that the marginal utility of the addictive good rises with the addictive state. This assumption is necessary (but not sufficient) to generate a property known as "adjacent complementarity," which means that greater current consumption leads to greater consumption in the future. According to Becker and Murphy, this is the distinguishing feature of an addictive substance.

Becker and Murphy's model generates a variety of interesting positive results regarding the use of addictive substances. For example, with appropriate parameterizations, the model generates behavior that is consistent with aspects of bingeing cycles and abrupt withdrawals. It distinguishes between conditions that lead to certain behaviors which they associate with addiction, and conditions that do not. It also predicts that an anticipated *future* increase in the

price of an addictive substance leads to an immediate decrease in drug use (see Gruber and Koszegi [2001] and Chaloupka and Warner [2001] for a review of supporting evidence).

From a normative perspective, the theory of rational addiction makes no distinction between addictive substances and other goods. Accordingly, the standard welfare theorems apply. It follows that government intervention is justified only if markets for addictive substances function imperfectly. There are two main concerns in this regard. First, if people are either poorly informed or misinformed about the effects of addictive substances, they may make poor decisions. As long as the government can provide relevant information more effectively and efficiently than private markets, educational policies (e.g., public health campaigns) are potentially beneficial. Second, the consumption of addictive substances may generate externalities. For example, driving under the influence leads to accidents, addicts commit crimes to support their habits, and addiction can be devastating to family members. The standard policy prescription for externalities involves a Pigouvian tax per-unit of the substance equal to the marginal external damage that it imposes on others.

Since the publication of Becker and Murphy's paper, others have extended the theory of rational addiction in a variety of ways, mainly to account for other observed features of addictive behavior. For example, in Orphanides and Zervos [1995], different people have different susceptibilities to addiction, which they discover through experimentation. The paper shows that a highly susceptible individual can control his addictive tendencies if he discovers his susceptibility quickly, but not if he discovers it slowly. The authors briefly discuss a few policy implications. Clearly, consumers benefit from accurate information concerning the distribution of susceptibilities. Moreover, since people are uncertain about their addictive susceptibilities, imperfections in private markets for rehabilitation insurance can leave them with residual risk, which potentially creates a role for government as a provider of social insurance. Other contributions include (but are not limited to) Dockner and Feichtinger [1993], who show how the theory of rational addiction can account for cyclical consumption patterns, and Orphanides and Zervos [1998], who introduce impulsiveness by allowing the consumer's discount rate to depend (in a time-consistent way) on use.

4.C. Some problematic empirical observations

In some ways, consumption patterns for addictive substances are no different than for other goods. A number of studies have shown that aggregate drug use responds both to prices and to information about the effects of addictive substances. For example, an aggressive U.S. public health campaign is widely credited with reductions in smoking rates. There is also

evidence that users engage in sophisticated forward-looking deliberation, reducing current consumption in response to anticipated price increases.[12] What, then, makes addiction a distinctive phenomenon? Bernheim and Rangel [2004] list five important behavioral patterns distilled from the extensive body of research on addiction in neuroscience, psychology, and clinical practice.

1. Unsuccessful attempts to quit. Addicts often express a desire to stop using a substance permanently and unconditionally but are unable to follow through. Short-term abstention is common while long-term recidivism rates are high. For example, during 2000, 70 percent of current smokers expressed a desire to quit completely and 41 percent stopped smoking for at least one day in an attempt to quit, but only 4.7 percent successfully abstained for more than three months.[13] This pattern is particularly striking because regular users initially experience painful withdrawal symptoms when they first attempt to quit, and these symptoms decline over time with successful abstention. Thus, recidivism often occurs after users have borne the most significant costs of quitting, sometimes following years of determined abstention.

2. Cue-triggered recidivism. Recidivism rates are especially high when addicts are exposed to cues related to past drug consumption. Long-term usage is considerably lower among those who experience significant changes of environment.[14] Treatment programs often advise recovering addicts to move to new locations and to avoid the places where previous consumption took place. Stress and "priming" (exposure to a small taste of the substance) have also been shown to trigger recidivism.[15]

3. Self-described mistakes. Addicts often describe past use as a mistake in a very strong sense: they think that they would have been better off in the past as well as the present had they acted differently. They recognize that they are likely to make similar errors in the future, and that this will undermine their desire to abstain. When they succumb to cravings, they sometimes characterize choices as mistakes even while in the act of consumption. It is instructive that the twelve-step program of Alcoholic Anonymous begins: "We admit we are powerless over alcohol - that our lives have become unmanageable."

As an example, Goldstein [2001,p.249] describes an addict who had been

---

[12] See Chaloupka and Warner [2001], MacCoun and Reuter [2001], and Gruber and Koszegi [2001] for a review of the evidence.

[13] See Trosclair et. al. [2002], Goldstein [2001], Hser, Anglin, and Powers [1993], Harris [1993], and O'Brien [1997].

[14] See Goldstein [2001], Goldstein and Kalant [1990], O'Brien [1976,1997], and Hser et. al. [1993,2001]. Robins [1974] and Robins et.al. [1974] found that Vietnam veterans who were addicted to heroin and/or opium at the end of the war experienced much lower relapse rates than other young male addicts during the same period. A plausible explanation is that veterans encountered fewer environmental triggers (familiar circumstances associated with drug use) upon returning to the U.S.

[15] See Goldstein [2001] and Robinson and Berridge [2003].

> "...suddenly overwhelmed by an irresistible craving, and he had rushed out of his house to find some heroin. ... it was as though he were driven by some external force he was powerless to resist, *even though he knew while it was happening that it was a disastrous course of action for him*" (italics added).

4. Self-control through precommitment. Recovering users often manage their tendency to make mistakes by voluntarily removing or degrading future options. They voluntarily admit themselves into "lock-up" rehabilitation facilities, often not to avoid cravings, but precisely because they expect to experience cravings and wish to control their actions. They also consume medications that either generate unpleasant side effects, or reduce pleasurable sensations, if the substance is subsequently consumed.[16] Severe addicts sometimes enlist others to assist with physical confinement to assure abstinence through the withdrawal process.

5. Self-control through behavioral and cognitive therapy. Recovering addicts attempt to minimize the probability of relapse through behavioral and cognitive therapies. Successful behavioral therapies teach cue-avoidance, often by encouraging the adoption of new life-styles and the development of new interests. Successful cognitive therapies teach cue-management, which entails refocusing attention on alternative consequences and objectives, often with the assistance of a mentor or trusted friend or through a meditative activity such as prayer. Notably, these therapeutic strategies affect addict's choices without providing new information.[17]

The clinical definition of addiction makes reference to some of these patterns. Substance addiction is said to occur when, after significant exposure, users find themselves engaging in compulsive, repeated, and unwanted use despite clearly harmful consequences, and often despite a strong desire to quit unconditionally (see e.g. the American Psychological Association's Diagnostic and Statistical Manual of Mental Disorders, known as DSM-IV).

From the perspective of traditional economic analysis, each of the patterns listed above is at least somewhat puzzling. The rational consumers of economic textbooks have no trouble following through on plans, and therefore should manifest neither of the first two patterns. Contrary to the third pattern, rational consumers always choose what they want, so, armed with good information, they can't make systematic mistakes. The notion that someone might be powerless over a consumption good is an anethema to a neoclassical economist. The standard

---

[16] Disulfiram interferes with the liver's ability to metabolize alcohol; as a result, ingestion of alcohol produces a highly unpleasant physical reaction for a period of time. Methadone, an agonist, activates the same opioid receptors as heroin, and thus produces a mild high, but has a slow-onset and a long-lasting effect, and it reduces the high produced by heroin. Naltrexone, an antagonist, blocks specific brain receptors, and thereby diminishes the high produced by opioids. All of these treatments reduce the frequency of relapse. See O'Brien [1997] and Goldstein [2001].

[17] Goldstein [2001] reports that there is a shared impression among the professional community that 12-step programs such as AA (p. 149) "are effective for many (if not most) alcohol addicts." However, given the nature of these programs, objective performance tests are not available. The AA treatment philosophy is based on "keeping it simple by putting the focus on not drinking, on attending meetings, and on reaching out to other alcoholics."

theory of consumer behavior embraces the principle that expanding or improving the set of available alternatives necessarily makes an individual better off, so precommitments can only be counterproductive, contrary to the fourth pattern. Finally, since in the standard model individuals never make mistakes, there is no role for expenditures on self-control.

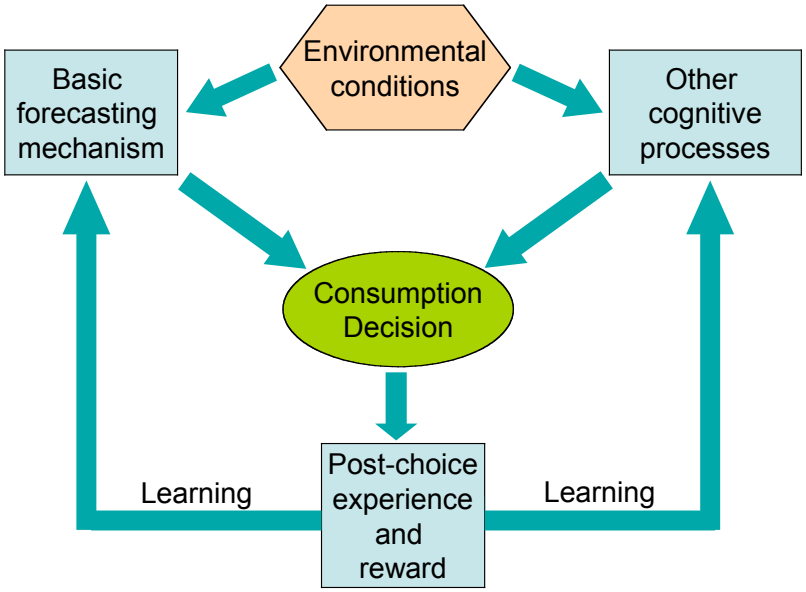
Creative extensions of the basic model may provide rationalizations for some of these patterns without overturning the basic paradigm. For example, Laibson [2001] has proposed a variant of the Becker-Murphy framework in which preferences become state-contingent with experience, and which can in principle account for cue management and avoidance. Even so, the five patterns described above collectively pose a serious challenge to neoclassical perspective, and provide motivation for economists to think "outside the box."

4.D. Recent insights from the neuroscience of addiction

Over the last 10 years, a new scientific consensus has begun to emerge concerning the nature of addiction. It now appears that addiction does not result primarily from the pleasurable effects of substances on the hedonic system. Instead, the new view of addiction holds that certain substances interfere with the proper operation of a neural system that plays an important role in learning. This is not to say that pleasure is unimportant. However, the key feature of addiction appears to be the fact that addictive substances cause a specific learning process to malfunction.

Figure 1 shows, at a high level of abstraction, how the brain normally makes decisions about standard consumption goods. Our senses provide us with information about environmental conditions. We process this information, along with information about our internal states -- things like hunger, fatigue, and so forth -- and this results in a decision. The decision is followed by experience, including rewards. The experienced relationship between environmental conditions, decisions, and rewards induces learning, which normally improves the quality of future decisions.

Figure 1: Decision Processes for Standard Consumption Goods

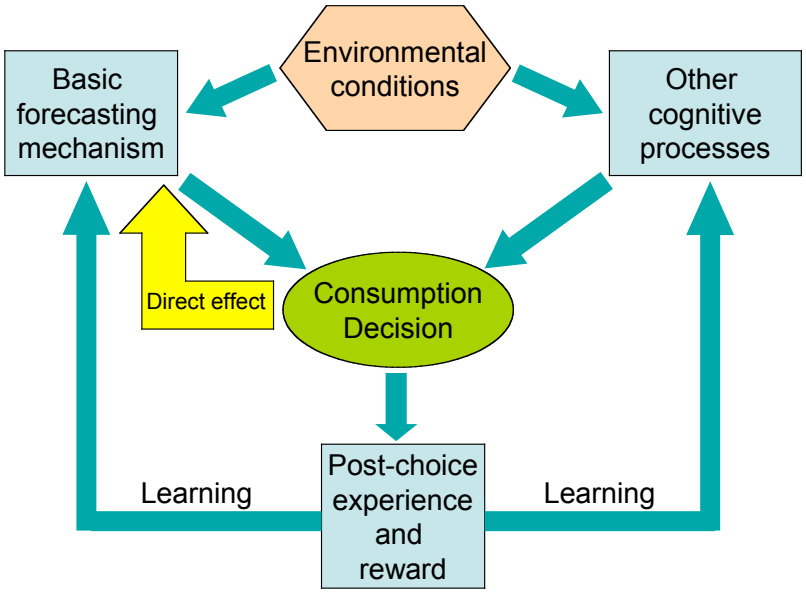


On left-hand side of this diagram, we've broken out an important component of the decision-making system, which we've labeled the "basic forecasting mechanism." This is a hard-wired system for measuring correlations between conditions, decisions, and short-term rewards. It does not involve higher reasoning; in fact, it's present in lower life forms as well as humans. For non-addictive substances, the basic forecasting mechanism learns with experience to construct an accurate forecast of the subsequent hedonic experiences.

It is worth emphasizing that the brain appears to have a variety of mechanisms for forecasting the possible consequences of decisions. Some involve higher cognition (represented on the right hand side of the diagram); for example, we sometimes develop causal models of the world and reason out the implications of our actions. Some – like the basic forecasting mechanism – are more mechanical.

Both types of forecasting mechanisms play a role in decision making. Sometimes we act based on the "gut reactions" generated by the basic forecasting mechanism. Sometimes higher cognition overrides a gut reaction. This is how the brain is designed to work. Each process has its advantages and disadvantages. The basic forecasting mechanism is very fast, but it's inflexible and unsophisticated. Higher cognition is flexible and sophisticated, but comparatively slow. When we have to make decisions quickly, we rely on our gut reactions. When there's no time pressure, we take the time to think things through. A balance between these systems emerged through evolution as nature's compromise. Consequently, the mere fact that we rely in some instances on impulses and gut reactions rather than reasoned deliberation does not mean that our

Figure 2: Decision Processes for Addictive Substances



choices are irrational or dysfunctional. For non-addictive substances, these mechanisms, operating in parallel, typically produce reasonable decisions.

Figure 2 shows how addictive substances interfere with the proper operation of these decision-making processes. In a nutshell, the problem with the addictive substances is that they act *directly* on the learning process underlying the basic forecasting mechanism, short-circuiting the neurological process by which this mechanism discovers correlations between environmental conditions, decisions, and rewards. As a result, the mechanism massively overstates the correlation between drug use and actual experienced pleasure. Loosely speaking, drugs fool a subconscious, hard-wired brain process into anticipating an exaggerated level of pleasure. An addict can try to compensate for this effect by exercising cognitive control, but he can't consciously correct the malfunction of the basic forecasting mechanism.

More specifically, the available neurological evidence supports four specific hypotheses that justify the new view of addictive substances (see Bernheim and Rangel [2004] for a more detailed discussion):

First, the mesolimbic dopamine system (MDS) serves, at least in part, as a basic forecasting mechanism which, with experience, learns to produce a response to situations and opportunities, the magnitude of which constitutes a forecast of near-term pleasure (see Schultz, Dayan, and Montague [1997] and Schultz [1998, 2000]).

Second, MDS forecasting does not appear to directly produce or reflect the experience of pleasure. Indeed, the human brain appears to contain a separate hedonic system that is responsible for producing sensations of "well-being." (see Berridge [1996,1999], Berridge and Robinson [1998,2003], and Robinson and Berridge [1993,2000,2003]).

Third, MDS-generated forecasts directly influence choices (see Berridge and Robinson [1998,2003] and Robinson and Berridge [1993,2000,2003])). In an organism with a sufficiently developed frontal cortex, higher cognitive mechanisms can override impulses resulting from MDS forecasts, for example by identifying alternative courses of action or projecting the future consequences of choices. The outcome depends on the intensity of the MDS forecast and on the ability of the frontal cortex to engage the necessary cognitive operations. A strong MDS forecast can impair this ability by influencing attention to stimuli, cognitive focus, and memory. Thus, a more attractive MDS-generated forecast makes cognitive override less likely.

We emphasize that the basic forecasting mechanism and higher cognitive processes are not two different sets of "preferences" or "selves" competing for control of decisions. Hedonic experiences are generated separately, and an individual maximizes the quality of these experiences by appropriately deploying both forecasting processes to anticipate outcomes.

Fourth, addictive substances act directly on the basic forecasting mechanism, disrupting its ability to construct accurate hedonic forecasts and exaggerating the anticipated hedonic benefits of consumption. Although addictive substances differ considerably in their chemical and psychological properties, there is a large and growing consensus in neuroscience that they share an ability to activate the firing of dopamine into the nucleus accumbens with much greater intensity and persistence than other substances. They do this either by activating the MDS directly, or by activating other networks that have a similar effect (see Nestler and Malenka [2004], Hyman and Malenka [2001], Nestler [2001], Wickelgreen [1997], and Robinson and Berridge [2003]). For non-addictive substances, the MDS learns to assign a hedonic forecast that bears some normal relation to the subsequent hedonic experience. For addictive substances, consumption activates dopamine firing directly, so the MDS learns to assign a hedonic forecast that is out of proportion to the subsequent hedonic experience. This not only creates a strong (and misleading) impulse to seek and use the substance, but also undermines the potential for cognitive override. Cognitive override still occurs, but in a limited range of circumstances.

The preceding discussion implies that, in some circumstances, drug use can literally be a mistake, in the sense that the brain is fooled into making a choice. It does not, however, imply that drug use is always a mistake. Even if the integrity of the basic forecasting mechanism is compromised, higher cognition can still either agree with it or override it. In different people,

brain chemistry appears to strike different balances between these mechanisms. This may explain why some people become addicts, while others use repeatedly without becoming addicted. Use can be rational in some instances and irrational in others. It is important to bear this point in mind when evaluating public policies alternatives.

In emphasizing the effects of addictive substances on decision processes, we do not mean to discount the significance of their hedonic effects. The typical user is initially drawn to an addictive substance because it produces a hedonic "high." Over time, regular use leads to hedonic and physical tolerance. That is, the drug loses its ability to produce a high unless the user abstains for a while, and any attempt to discontinue the drug may have unpleasant side effects (withdrawal). Cue-conditioned "cravings" may have hedonic implications as well as non-hedonic causes. All of these effects are clearly important. However, there is an emerging consensus in neuroscience and psychology that decision-process effects, rather than hedonic effects, provide the key to understanding addictive behavior (see Wise [1989], Robbins and Everitt [1996], Di Chiara [1999], Kelley [1999], Nestler and Malenka [2004], Hyman and Malenka [2001], Berridge and Robinson [2003], Robinson and Berridge [2000], and Redish [2004]).

4.E. Modeling addiction as a decision-process malfunction

Bernheim and Rangel [2004] present a theory of addiction that departs from the fourth assumption discussed in Section 2 (that choices are always aligned with preferences). The theory is based on the following three main premises.

First, use among addicts is sometimes a mistake, in the sense that actions diverge from preferences, and sometimes rational.

Second, experience with an addictive substance sensitizes an individual to environmental cues that trigger mistaken usage.

Third, addicts understand their susceptibility to cue-triggered mistakes and attempt to manage the process with some degree of sophistication.

The first two premises are justified by the body of research described in section 3.D, which shows that, after repeated exposure to an addictive substance, the brain tends to make skewed hedonic forecasts upon encountering environmental cues that are associated with past substance use. The third premise is justified by behavioral evidence indicating that users are often surprisingly sophisticated and forward looking. For example, they reduce current consumption in response to expected future price increases (Gruber and Koszegi [2001]). Some also enter detox not because they intend to remain sober, but rather because they want to increase the intensity of the next high.

A summary of the model. The formal model in Bernheim and Rangel [2004] envisions an individual who makes a sequence of decisions regarding lifestyle, the use of an addictive substance, and the consumption of non-addictive substances. It assumes that, at any point in time, the individual operates in one of two modes: a "cold" mode in which properly functioning decision-making processes lead to the selection of his most preferred alternative, and a dysfunctional "hot" mode in which decisions and preferences may diverge (because he responds to distorted MDS-generated forecasts).[18] The hot mode is transient, but always results in use of the substance. The likelihood of entering the hot mode at any moment depends on the individual's history of substance use, his chosen lifestyle (e.g., partying exposes the individual to more intense substance-related cues), and random events (e.g., the frequency and intensity of recently encountered environmental cues to which he has been sensitized through prior use).

The history of use is summarized through the notion of an addictive state. Use moves the individual to a higher addictive state, and abstention moves him to a lower addictive state. An increase in the addictive state raises the likelihood of entering the hot mode at any moment (e.g., because it implies increased sensitivity to randomly occurring environment cues). Higher addictive states are also associated with lower baseline well-being (e.g., due to deteriorating health), lower financial resources (due to decreased productivity, absenteeism, and out-of-pocket medical expenses), and possible a greater "boost" from consuming the addictive substance.

By varying assumptions about the properties of the substance in question, the model can replicate a wide range of observed behaviors. In particular, it can account for each of the patterns discussed in Section 3.C (see Bernheim and Rangel [2004] for details and Bernheim and Rangel [2005c] for simulations of the model).

Policy implications. This theory admits two classes of rationales for government intervention. First, as in the theory of rational addiction, intervention may be justified to correct market failures involving addictive substances – that is, the government can address externalities, misinformation, and ignorance. Second, policies may also affect the frequency and consequences of mistakes. This consideration gives rise to a number of non-standard policy implications.

1. Limitations of informational policy. In practice, public education campaigns (such as the U.S. anti-smoking and anti-drug initiatives) have achieved mixed results. The process-malfunction theory of addiction highlights a fundamental limitation of informational policy:

[18] Our analysis is related to work by Loewenstein [1996, 1999], who considers simple models in which an individual can operate either in a hot or cold decision-making mode. Notably, Loewenstein's approach relaxes the assumption of fixed life-time preferences. He assumes that behavior in the hot mode reflects the application of a "false" utility function, rather than a breakdown of the processes by which a utility function is maximized. He also argues, contrary to our findings, that imperfect self-understanding is necessary for addiction-like behaviors.

contrary to standard theory, one cannot assume that even a highly knowledgeable addict always makes informed choices. Information about the consequences of substance abuse may affect initial experimentation with drugs, but cannot alter the neurological mechanisms through which addictive substances subvert deliberative decision making.

2. Counterproductive disincentives. Policies such as "sin taxes" and criminalization strive to discourage use by making substances costly. As we've noted, this is potentially justifiable on the grounds that use generates negative externalities. In the context of the theory described in this section, even higher taxes (whether implicit or explicit) might be justified if they reduce excessive use in "hot" decision states. Unfortunately, it is likely that compulsive use of addictive substances is much less sensitive to costs and consequences than is deliberative use. Consequently, imposing costs in excess of external diseconomies is likely to distort cold-state choices detrimentally, without significantly reducing problematic hot-state usage. Indeed, policies that impose high costs on use may thwart social insurance objectives by exacerbating the consequences of uninsurable risks associated with the use of addictive substances.[19] Accordingly, the optimal rate of taxation for addictive substances may be significant *lower* than that implied by externalities (see Bernheim and Rangel [2005c] for simulation results).[20]

3. Supply disruption. Standard reasoning suggests that taxation is preferable to criminalization. Both impose costs, but taxes generate revenues, while criminalization dissipates social resources. In the context of the theory discussed in this section, criminalization offers an offsetting benefit: it disrupts supply, making it particularly difficult for users to obtain a banned substance on short notice. The effect on use is likely to be larger in hot states, when people act impulsively, than in cold states, when people plan deliberatively. This is exactly what one would hope to achieve, and precisely opposite the effect of a tax. To put it somewhat differently, criminalization may help some addicts impose self-control, without (as a practical matter) preventing deliberate use. There is, however, an associated disadvantage: while in the hot state, addicts may engage in costly and potentially dangerous search.

4. Beneficial harm reduction. If addiction results in significant part from randomly occurring mistakes, various interventions can serve social insurance objectives by ameliorating some of its worst consequences. For instance, subsidization of rehabilitation centers and treatment programs (particularly for the indigent) can moderate the financial impact of addiction and promote recovery. Likewise, the free distribution of clean needles can moderate the incidence of diseases among heroin addicts. In some cases, it may even be beneficial to make substances

[19] In practice, addicts often suffer severe economic deprivation, turning to crime and prostitution for support. High substance costs aggravate these consequences.
[20] As shown in Bernheim and Rangel [2004], this result depends on usage patterns.

available to severe addicts at low cost.[21] As is usually the case, one must trade off the benefits from insurance against incentive effects: by moderating consequences, harm-reducing policies could in principle encourage casual use and experimentation.

5. Policies affecting cues. Since environmental cues frequently trigger addictive behaviors, public policy can also influence use by changing the cues that people normally encounter. One approach involves the elimination of problematic cues. For example, advertising and marketing restrictions of the type imposed on sellers of tobacco and alcohol suppresses one possible artificial trigger for compulsive use. Since one person's decision to smoke may trigger another, confining use to designated areas may reduce unintended use. A second approach involves the creation of counter-cues. For example, Brazil and Canada require every pack of cigarettes to display a prominent, viscerally charged image depicting some deleterious consequences of smoking, such as lung disease and neonatal morbidity. In principle, a sufficiently strong counter-cue could trigger thought processes that induce users to resist cravings, even though the same information is ineffective when offered in a less provocative format. Policies that eliminate problematic cues or promote counter-cues are potentially beneficial because they combat compulsive use while imposing a minimal inconvenience and restrictions on deliberate rational users.

6. Facilitation of self-control. The process-malfunction theory of addiction places a high value on policies that provide better opportunities for self-regulation without making particular choices compulsory. This potentially helps those who are vulnerable to compulsive use, without encroaching on the freedoms of those who would deliberately choose to use. Laws that limit the sale of a substance to particular times, places, and circumstances frequently provide limited opportunities along these lines (see e.g. Ornstein and Hanssens [1985], Norstrom and Skog [2005], and Tigerstedt and Sutton [2000]). Well-designed policies could in principle accomplish this objective more effectively. For example, a number of states have enacted laws allowing problem gamblers to voluntarily ban themselves from casinos (Yerak [2001]). Alternatively, if a substance is available only by prescription, and if prescription orders are filled on a "next day" basis, then deliberate forward-looking planning becomes a prerequisite for availability. Recovering heroin addicts could self-regulate problematic compulsive use by carefully choosing when, and when not, to file requests for refills.

[21] For example, Swiss policy makes heroin available at low cost to severe addicts.

4.F. Modeling addiction with quasi-hyperbolic discounting

One important line of work modifies Becker and Murphy's model of "rational addiction" by adding quasi-hyperbolic (β,δ)-discounting (see Gruber and Koszegi [2001, 2004] and O'Donoghue and Rabin [2000]).[22] In contrast to the theory of rational addiction, the consumer acts as if he attaches disproportionate importance (1/β) to current well-being when making decisions about current consumption.

Gruber and Koszegi use this model to compute optimal cigarette taxes. When evaluating individual welfare, they assume that true preferences correspond to standard exponential discounting. Implicitly, they adopt the interpretation of quasi-hyperbolic discounting discussed in Section 2.D: true preferences are standard, but the decision-making process leads individuals to make present-biased mistakes, which the (β,δ)-model captures in reduced form.

In principle, one could defend this interpretation with reference to the evidence described in Section 2.D. Unfortunately, the model does not fit these facts in two important respects. First, the evidence indicates that mistakes are domain-specific. In contrast, the proclivity to make present-biased mistakes in the (β,δ)-model cuts across all domains. Second, the evidence indicates that mistakes are triggered by intermittent environmental cues. In contrast, the decision maker *always* suffers from present-bias in Gruber and Koszegi's framework.

One could, of course, formulate a variant of Gruber and Koszegi's model with narrow-domain, cue-triggered present-bias. The resulting model would be a close cousin of the process-malfunction theory of addiction discussed in the previous section. However, one significant difference would remain. In the β-δ framework with the proposed modifications, the decision maker would remain sophisticated, forward-looking, and responsive to economic incentives even when suffering from present-bias. In contrast, the process-malfunction theory holds that mistakes result from simple impulses generated by a hard-wired process that encompasses a limited range of consequences.

In some respects, the policy implications of this approach are similar to those discussed in the preceding section. Informational policy alone is limited because it cannot address the causes of present-bias. Supply disruption is potentially beneficial, as are policies that facilitate the exercise of self-control.

In other respects, the policy implications described by Gruber and Koszegi differ sharply from those discussed in the preceding section. Most notably, the β-δ framework provides a rationale for "sin taxes" (see also O'Donoghue and Rabin [2005]). When making decisions, the

[22] In an earlier related paper, Winston [1980] modeled addiction by assuming that lifetime preferences vary with states of nature.

consumer always puts too little weight on future consequences, including those resulting from adjacent complementarities. The government can address these "internalities" (externalities imposed on future selves) by imposing a Pigouvian tax on current consumption. Accordingly, the rate of taxation for addictive substances should be higher than that justified by marginal externalities. For example, according to Gruber and Koszegi's simulations, the optimal tax on each pack of cigarettes is at least a dollar higher than would be justified by externalities alone.

Why do the models of Bernheim-Rangel and Gruber-Koszegi lead to sharply differing conclusions concerning substance taxation? The answer lies in two of the issues discussed above. First, Gruber and Koszegi assume that consumers *always* make present-biased mistakes, while Bernheim and Rangel assume that mistakes occur only in the presence of intermittent environmental cues. Accordingly, social insurance can enhance the consumer's well-being in Bernheim-Rangel, but not in Gruber-Koszegi. In other words, Gruber and Koszegi's assumptions eliminate the factor that argues *against* high tax burdens in Bernheim and Rangel's model. Second, Gruber and Koszegi assume that the decision maker remains sophisticated, forward-looking, and responsive to economic incentives even while committing errors, whereas Bernheim and Rangel assume that errors result from a mechanical and largely inflexible process. Accordingly, taxation directly reduces decision errors in Gruber-Koszegi, but has a limited effect along these lines in Bernheim-Rangel.

4.G. Modeling addiction with temptation preferences

Gul and Pesendorfer [2005] propose a model of addiction based on the temptation preferences discussed in section 3.G. Following their earlier work on temptation (Gul and Pesendorfer [2001]), they assume that the consumer's preferences are defined both over consumption bundles and over the sets from which these bundles are chosen. In each period of life, the consumer divides his resources between two goods, one of which is addictive, with the object of maximizing an intertemporal utility function. This function is standard in all respects, except that it is modified to include, for each period, a penalty representing net temptation from the most tempting unchosen alternative in the choice set. Even though the consumer applies the same lifetime preferences at every moment in time and makes no mistakes, precommitments are still potentially valuable because they reduce the unpleasant feelings associated with the temptation to consume addictive substances.

Gul and Pesendorfer's model invokes a number of important assumptions. The following three deserve emphasis. First, the level of temptation associated with an alternative depends only on the level of the addictive good, and not at all on the level of the non-addictive good. Second,

recent consumption of the addictive substance increases the weight given to temptation, but does not enter the "standard" portion of the utility function. According to this assumption, as long as an individual is forced to abstain from the addictive substance, his experienced well-being is unrelated to his past consumption. As a result, this assumption is in sharp conflict with evidence on cravings and withdrawal. Third, the consumer only experiences temptation with respect to current choices. For example, when deciding whether to enter rehabilitation for the next period, he is not tempted by the prospect of future drug use.

In the Gul-Pesendorfer model, private markets tend to work poorly relative to the first-best. Markets provide people with choices, and choices create costly temptation. Unless it is possible to irrevocably lock in all choices in advance, a consumer is typically happier with the first-best consumption trajectory when someone else chooses it for him, than when he chooses it himself in "real time" (it is first-best in the first instance, but not in the second).

Even though the laissez faire solution is inefficient, the optimal rate of taxation or subsidization for an addictive good is zero. The same result holds in standard models of commodity taxation (when the government has no revenue requirement), for essentially the same reasons. However, we conjecture that this is a knife-edge case, driven by the first assumption mentioned above. It would appear that if, contrary to the assumption, temptation depends, at least to some extent, on immediate rewards from the non-addictive good (in addition to consumption of the addictive substance), the optimal rate of sin taxation is strictly positive.[23]

Other policy implications resemble those discussed in previous sections. Informational policy alone is limited because it cannot address the causes of temptation. Supply disruption is potentially beneficial because it removes tempting alternatives. Policies that facilitate self-control can also enhance welfare by allowing consumers to eliminate alternatives that would otherwise prove tempting in the future.

4.H. Looking Ahead

The case of addiction exemplifies the potential for improving policy analysis through the integration of psychology, neuroscience, and economics. Though progress is evident, much work remains. We close this section with a brief discussion of some important open questions.

[23] Holding the consumption level for the addictive substance fixed, an increase in the rate of sin taxation reduces the consumption level of the non-addictive good, rendering the alternative less attractive. Since the size of this effect is proportional to the quantity of the addictive substance, taxation presumably reduces the "temptation gap" between alternatives with low and high levels of addictive consumption. Furthermore, this is a first-order effect. Accordingly, one suspects, intuitively, that a small positive tax is welfare-enhancing. We have not yet attempted to verify this conjecture formally.

Estimation and testing of competing behavioral models. Almost all of the existing empirical work on addictive behavior is either atheoretical (i.e., it documents factual patterns) or based on the framework of rational addiction. So far, research on behavioral alternatives has been almost exclusively theoretical. It is important to explore the feasibility of estimating parsimonious structural versions of the various competing behavioral models, using both choice data and a combination of choice and non-choice data. Insights from ongoing research in neuroscience should be exploited to develop procedures for acquiring and using new types of pertinent non-choice data (e.g., on physical states). Future research should compare the performance of the models in explaining observed behavior, and examine testable implications that distinguish between them. Empirical research can potentially shed light on the relative importance of the various forces at work in these models.

Imperfect foresight. Most of the economic literature on addiction assumes that people perfectly understand the benefits and costs of substance use, including its effects on future tastes and decision-making processes. The evidence suggests that this extreme assumption is unrealistic. For example, in a study of high-school seniors who smoked cigarettes, 56% predicted that they would not be smoking in 5 years, but in fact only 31% were able to quit (USDHH [1994]).

Under the assumption that decision makers are completely or partially naïve, models with quasi-hyperbolic discounting incorporate imperfect self-understanding. While this represents a step in the right direction, further work is clearly needed. Models of naïve behavior should draw on new and existing empirical research concerning the nature of unsophisticated decision making. They should allow for the possibility that people lack perfect foresight not only with respect to their own future tastes and choices, but also with respect to other consequences, such as health effects (e.g., as in Hung [2000]). They should also introduce the possibility that people learn about their self-control problems with experience.

The literature on "projection bias" (e.g., Loewenstein, O'Donoghue, and Rabin [2003]) illustrates the potential to discover important regularities concerning the structure of naïve decision making through empirical research. This phrase refers to the tendency for people to assume that their future likes and dislikes will be more similar to their current likes and dislikes than is actually the case.[24] Loewenstein, O'Donoghue, and Rabin [2003] briefly and informally discuss several provocative implications for addiction. Victims of projection bias are more likely

[24] Projection bias does *not* imply that *lifetime* preferences vary from one point in time to another. On the contrary, an otherwise standard consumer suffering from projection bias wants future tastes to govern future choices. However, he makes decisions based on biased forecasts of future tastes.

to become addicted against their interests because they underestimate both the effects of habit formation and the degree to which current consumption has negative consequences for future health. Once addicted, they are more likely to try to quit when they are not experiencing cravings, because they underestimate future cravings. Conditional on attempting to quit, they are also more likely to "fall off the wagon" because, upon experiencing cravings, they overestimate the difficulty of continued abstention in the future.

Differences across substances and populations. It is important to emphasize that there is no single combination of policies that is ideal for all addictive substances. For example, while alcohol and crack cocaine are both addictive, public policy should (and does) treat them differently. A number of factors affect the relative desirability of the various policy alternatives, including (but not limited to) the typical individual's susceptibility to addiction, the responsiveness of compulsive and deliberative use to prices and other incentives, and the magnitude of the externalities imposed on third parties. It is also important to stress that the ideal policy regime for any particular substance may evolve over time as our ability to treat, control, and/or predict addiction develops. Ideally, economists should attempt to estimate parametric behavioral models for a wide range of substances and populations, and to use these estimates as a basis for determining the best policy for each substance.

5. Public Goods

In this section, we review the contributions of behavioral public economics to our understanding of public goods. As in previous sections, we identify the key policy issues, summarize the standard approach, and discuss empirical evidence that calls this approach into question. We then review the leading behavioral alternative and discuss its implications.

5.A. The policy problem

A large number and wide variety of public policy issues –from the environment to school finance, and from the war on poverty to the financing of basic research – involve the provision of public goods. Funding for these goods flows from both public and private sources. At the community level, philanthropic activities in the U.S. address a large class of socially valuable activities, from assisting the poor to financing cultural events. Andreoni [2004] reports that, for the U.S., contributions to the philanthropic sector totaled 240.3 billion dollars in 2003; moreover, in 1997, roughly 45,000 charitable, religious, and other non-for-profit organizations were

registered with the government.[25] Voluntarily provided public goods also play important roles in smaller groups, such as families.

In each of these domains, public goods give rise to a common problem: how can the group best overcome free riding and provide funding at an appropriate level? Should the group provide its members with incentives to contribute (e.g., tax breaks)? Should it require mandatory contributions (e.g,, through taxes)? Is it best to have a hybrid system that draws on both public and private contributions?

To answer these questions, economists require a theory of public goods that explains observed patterns of voluntary giving. The theory must explain why people give, how they select the causes to which they contribute, and how their contributions respond to economic variables, government policies, and the behavior of others. It should also account for the existence of philanthropic organizations, and explain how the activities of these entities respond to government policy.

5.B. The neoclassical perspective on public goods

The standard model of public goods assumes that each member of a group of N individuals has true preferences over consumption of private goods (denoted x^i) and public goods (denoted G). These preferences are represented by a utility function $U^i(x^i,G)$. For expositional simplicity, we focus here on a simple model with only one private good and one public good, where one unit of the private good is required to produce each unit of the public good, and where each individual i is endowed with w^i units of the private good. All of the results described below generalize to more complicated settings. Each individual contributes an amount g^i to the public good. In addition, individual i pays a lump-sum tax, T^i, and the government contributes all revenues to the public good. Consequently, $G = (g^1 + T^1) + \ldots + (g^N + T^N)$. Individuals simultaneously select their contributions after learning the values of the lump-sum taxes. Behavior is governed by Nash equilibrium. Let g^{i*} denote the equilibrium level of contributions, and $G^* = g^{1*} + \ldots + g^{N*}$ denote the equilibrium level of public goods.

It is useful to highlight the key assumptions built into this framework. First, individuals only care about their consumption of private and public goods. They do not benefit *directly* from making contributions, nor do they care about others' consumption or well-being. Second, individuals do not care about the *process* through which allocations are determined. For example, they are indifferent between public and private provision as long as the level of private

[25] The sources of these funds are as follows: 76.3% came from individuals, 11.2% came from foundations, 7.5% from bequests, and the remaining 5.1% was given by corporations.

consumption and public good provision is the same in both instances. Notice also that, in this simple model, there is no obvious role for charitable fundraising. For example, since people are fully informed about the public good, there is no reason for charities to disseminate information.

This model has featured prominently in several important strands of the literature. These include work on optimal tax and regulatory policy in the presence of externalities, the design of efficient mechanisms for public goods problems , and political economy models of public goods provision. From a positive perspective, the model has a number of sharp, testable implications, including the following (see Bergstrom, Blume, and Varian [1986] and Andreoni [1988] for details):

1.Extreme income elasticities. If individuals have identical preferences, there exists an endowment level w^* such that (a) only those with an endowment larger than w^* contribute, and (b) $g^{i*} = w^i - w^*$ otherwise. The result extends to the case of heterogeneous tastes as long as each taste-type is represented across the income distribution. It follows that the marginal propensity to contribute to the public good is exactly unity (measured in the cross-section, controlling for individual characteristics) for those with sufficiently high resources,[26] and exactly zero for the rest of the population. It also follows that all contributors (of the same type) consume the same amount of private goods.

2. Only the wealthy contribute. In large groups only the very upper tail of the income distribution contributes to the public good. Furthermore, as the population grows (fixing the distribution of wealth), contributors account for a smaller fraction of the population. As a result, the effect of population size on total contributions converges to zero for sufficiently large populations. Unless the group is small, the level of public goods depends only on the wealth of the very rich: changes in wealth for the rest of the population have no impact on total provision.

3. Neutrality of public provision. Public provision of public goods financed through lump-sum taxation is neutral as long as no individual pays a lump-sum tax greater than the contribution he would make in the absence of government intervention. In this case public contributions fully crowd out of private contributions. While the conditions required for neutrality seem stark, the result generalizes to other environments. For example, Bernheim [1986] and Andreoni [1988] have shown that the total level of the public goods is invariant, or approximately invariant, with respect to public provision financed by distortionary taxes, and with respect to subsidized giving. These results build on earlier work by Warr [1982] and Roberts [1984].

[26] In response to an exogenous increase in resources (as opposed to cross-sectional variation), a contributor will increase private consumption. However, if the number of contributors is large, the recipeint's marginal propensity to consume the private good is approximately zero.

4.Contributions from external sources (almost) fully crowd out internal funding. In a large economy, exogenous contributions to the public good (made by someone outside the group, say a higher level of government) have a negligible impact on the level of provision. In other words, external funding almost fully crowds out of private contributions. It follows, for example, that contributions from a higher level of government to a local charity cannot measurably increase total funding, assuming the number of contributors is reasonably large.

5. Neutrality of redistribution. Redistributing wealth among contributors has no effect on the total level of contributions. In contrast, redistributing wealth from the group of contributors to the group of non-contributors decreases the total level of the public good.

These results are valuable because they provide stark and robust testable implications of the standard model. How well do they match the data?

5.C. Some problematic observations

One of the most influential empirical tests of the standard model is Kingma [1989]. In contrast to the bulk of the literature that preceded it, this paper studies contributions to a *particular* public good – the operation of public radio stations – rather than aggregate contributions. The narrow focus is desirable because, when analyzing aggregates, it is difficult to harmonize the scope of data pertaining to public and private contributions. Moreover, a high rate of giving in the aggregate may mask low rates of giving to individual causes. The paper uses a unique cross-sectional dataset on the funding sources and member contributions to 66 public radio stations across the U.S. serving non-overlapping markets. It has two main findings. First, about half of the subjects in the sample (who were recruited for a study of listening habits) contribute positive amounts. The average contribution given was $45. Contributors were wealthier and more educated on average, but not by a significant amount. This finding stands in sharp contrast to implications 1 and 2 from the previous section. Second, a $10,000 increase in "exogenous" public contributions to the station (that is, contributions financed by federal taxes rather than taxes on local members) reduces private contributions by $1,350 for a typical station with 9,000 members. This contradicts implication 4.

Kingma's first finding is consistent with patterns observed in the aggregate data. For example, Andreoni [2004] reports that, in 1995, 68.5% of all households gave to charity, and the average gift amount was $1081. Even relatively poor households gave almost 5% of their incomes, on average, to charity; as a fraction of income, households in upper-income brackets actually gave less.

Kingma's second finding is also roughly consistent with other studies based on aggregate data. For example, Abrams and Schmitz [1978a,b] and Clotfelter [1985] find that public transfers to the 'non-for-profit' sector crowd out private giving at the rate of 5 to 28 cents on the dollar.

The first four implications listed in the previous section have also been tested in the laboratory. Isaac and Walker [1988] study the effect of group size in linear public good experiments. Subjects play repeatedly with either 3 or 9 other participants. Each round they receive an endowment of tokens and decide how many tokens to contribute to the public good. Tokens are valuable because they are exchangeable for cash at the end of the experiment. Each token contributed to the public good yields either 0.3 or 0.7 tokens for everyone in the group, including the contributor. Since each token contributed entails a net loss, the standard model predicts that it is a dominant strategy for every subject to contribute nothing. As in many other experiments in this literature, subjects initially contribute roughly 50% of tokens on average, but this figure falls as the experiment is repeated. Neither average individual contributions nor the fraction of subjects contributing a positive amount decline with group size. These findings contradict implication 2.

Andreoni [1993] studies a variant of the previous experiment in which payoffs vary non-linearly with the number of tokens. This generates a Nash equilibrium with strictly positive contributions. He tests the neutrality of public provision (implication 3) by comparing behavior in two closely related treatments. In each case, subjects choose how many tokens to contribute and are given a 2-dimensional table that describes how their payoffs change as a function of their contribution and the aggregate contributions of others. In one treatment, they are, in effect, required to contribute at least two tokens; in the other treatment, they are not required to contribute anything.[27] Andreoni's results imply that public contributions crowd out private contributions at the rate of 71 cents on the dollar. While this rate of crowding-out is high in comparison with other estimates in the literature, it is still inconsistent with implication 3.

These papers, together with a growing body of related evidence (see Ledyard [1995] and Camerer [2003] for reviews), have lead many economists to reject the standard model, and to search for superior alternatives. The rest of this section reviews the state of the literature and summarizes its implications for public economics.

5.D. Models involving "warm glow"

[27] Given the importance of framing effects in social exchange experiments, it is noteworthy that the minimum contribution level is imposed by restating the payoffs associated with a given contribution profile, rather than by retaining the same payoff mapping and restricting choices.

To account for the evidence described in the preceding section, Andreoni [1989,1990] proposed a "warm-glow" model of public good contributions, which builds on ideas in earlier papers by Blinder [1974], Becker [1974], Cornes and Sandler [1984] and Steinberg [1987]. His approach entails a straightforward modification of the standard model: individuals are assumed to behave as if they maximize a utility function of the form $U^i(x^i, g^i, G)$ instead of $U^i(x^i,G)$. In this formulation, each individual cares *directly* about the amount he contributes to the public good, in addition to his consumption of the private and public goods.

This modification overturns each of the implications discussed in the preceding section, and leads to more sensible policy implications. For example, as the size of the population increases, choosing a contribution level becomes more and more like picking the level of consumption for any conventional good. In the limit, the contributor simply weighs the relative merits of spending money on two different private goods, x^i and g^i; the effect on his well-being through G becomes negligible Accordingly, the model can produce sensible income elasticities and high rates of charitable giving throughout the income distribution. The level of the public good is responsive to changes in the income distribution, public provision increases funding levels whether financed by taxes on group members or by external sources, and redistributions among contributors are non-neutral. In fact, in the warm-glow model, the optimal tax treatment of charitable contributions qualitatively resembles the U.S. tax code (Diamond [2005]). In short, the implications of the warm-glow model are more consistent than the standard framework both with the empirical findings described in the previous section, and with the perspective of policy makers.

In contrast to some of the work on addiction or saving summarized in the last two sections, the literature on warm-glow giving has had little to say about the mechanisms responsible for generating departures from the standard framework. While it is plainly appropriate to think of the model as a reduced form representation of a more complex underlying process, the nature of this process is largely unexplained.

A partial list of possible warm-glow mechanisms includes the following. First, people may experience positive emotions (e.g., pride) when they conform to or exceed certain standards of "virtuous" behavior, or negative emotions (e.g., guilt) when they fall short of these standards. Second, they may be concerned about the inferences that others draw from their actions (for example, whether they are generous or public-spirited), and this may increase their willingness to contribute (Harbaugh [1998], Shang and Croson [2005]). Third, upon forming a group, people may contribute to establish a norm of positive reciprocity, thereby promoting future cooperation. Fourth, when it is possible for group members to inflict harm on each other, giving may rise in

response to implicit or explicit threats (negative reciprocity) that become credible as a result of emotional responses, such as anger (see Fehr and Gachter [2000,2002], Ferhr and Fischbacher [2003,2004], Sefton et. al. [2002] and Masclet et. al. [2003]).

One of the main themes of this paper is that a good understanding of pertinent psychological and neural processes is often helpful in formulating reduced form models that can faithfully reproduce observed patterns and reliably predict behavior out of sample, as well as in justifying specific normative criteria. Unfortunately, in the context of warm glow giving and public goods, these processes are not yet well understood. The warm glow model remains a "black box," and one can interpret it as a reduced form for a variety of mechanisms with starkly differing welfare implications. Diamond [2005] argues that, given the limited state of knowledge concerning processes, measures of social welfare should exclude the apparent benefits from the warm glow. He advocates using the warm-glow model for positive purposes (that is, to describe behavior), but favors the standard model for evaluating welfare. Andreoni [2004] expresses a similar view, and in addition argues that economists are unlikely to shed much light on the nature of the true preferences that give rise to warm-glow behavior. While we are more sanguine about the prospects for meaningful progress, we agree that economists do not yet understand warm-glow mechanisms sufficiently well to resolve important questions about positive and normative analysis.

One concern is that apparent warm glow behavior may sometimes reflect a divergence between choices and true preferences. In some instances, people may give because they derive pleasure from the act. For example, giving to a worthy cause may make them feel proud to have taken constructive action, or it may assuage their guilt. In such cases, revealed preference provides a reasonable basis for welfare evaluation (subject to the further qualifications discussed below). However, in other situations, exposure to an emotionally manipulative message may precipitate giving by triggering a short-lived emotional reaction such as shame, and people may experience remorse shortly thereafter (see Loewenstein [1996] and Loewenstein and O'Donoghue [2004] for other interesting examples of this type of phenomena). Different implications for welfare follow depending on whether the individual, when in a normal state of mind, wishes to limit his ability to give upon encountering an emotional trigger. If behavior is dynamically consistent, it may be appropriate to adopt a state-contingent version of the warm-glow model for both positive and normative analysis. However, if behavior is dynamically inconsistent, it may be appropriate to discount the "revealed" impact of giving on transient perceptions of well-being (e.g., for the reasons discussed at the end of Section 2.C).

A second concern is that warm glow effects appear to be context-dependent. Experiments have shown that the amount of giving depends on framing, the identity of the group, the emotional state of the subjects, and the history of play. Here we mention two examples. First, Andreoni [1995] finds that a change in the phrasing of instructions can have a sizable effect on contributions, even when strategy sets and payoffs are unchanged. Second, Isaac and Walker [1988] (and dozens of subsequent studies) find that the level of contributions decreases with repetition in both small and large groups.[28] Furthermore, the rate of decline depends on the behavior of others: subjects are more likely to stop acting cooperatively if others behave selfishly (see Ledyard [1986] and Camerer [2003] for surveys of the literature).

This concern is relevant for policy analysis. The appropriate reduced-form representation of warm-glow giving may vary from one set of policies and institutions to another. The practice of forecasting the behavioral effects of a policy change based on fixed warm-glow parameters is therefore vulnerable to the Lucas critique. For example, if people experience less pride when making contributions in the presence of economic incentives, subsidization of contributions will stimulate less giving than anticipated, and could even reduce it. Alternatively, people may become less resistant to taxation if they are regularly supplied with more concrete and visual evidence of the benefits derived from public expenditures. Public relation campaigns that show "your tax dollars at work" are, in effect, intended to foster a warm glow.

One could, of course, modify the warm-glow model to account for context-dependence by linking the taste for giving to features of the environment in just the right way. However, this solution is conceptually unsatisfactory. One cannot usefully explain a phenomenon by selecting ad hoc preferences that rationalize choices ex post (Stigler and Becker [1977]). If every context is potentially associated with a different mode of behavior, out-of-sample prediction is impossible. To anticipate the positive effects of policy changes, one therefore needs a broad theory that accounts for the relevance of context. This requires us to open the black box.

Since the warm-glow mechanism may differ from one context to another, a deeper understanding of context-dependence is also essential for welfare analysis. To illustrate the problem, consider the following hypothetical example. How should we evaluate a policy that replaces private contributions to a public good with tax-financed contributions, without changing either the total amount obtained from any individual or the overall level of funding? Are people worse off because they lose the beneficial warm glow associated with giving? Are they better off

[28] Palfrey and Prisbey [1997] argue that the implied decline in the warm-glow taste parameter may in part reflect falling rates of decision errors. To our knowledge, there is no evidence that distinguishes between the hypothesis that tastes evolve with repetition and the possibility that error rates decline. However, errors do not appear to explain many other findings in this literature.

because public funding relieves them of guilt? Or are they equally well off because they experience the same warm glow from giving voluntarily and from paying their taxes? While it may be difficult to resolve this issue, we are optimistic about the prospects for progress through further research involving a combination of psychology, neuroscience, and experimental economics.

Despite these concerns, the theoretical and empirical literatures concerning warm-glow giving have already contributed significantly to our understanding of public goods, and have changed the way many public economists think about related policies. We can now say with some confidence that people act *as if* they care about the levels of their own contributions. We know that the intensity of this effect depends on context. While we lack a good theory of context-dependence, we have a good set of empirical regularities from which to build. We have good reason to believe that people feel differently about public and private contributions. We have both direct and indirect evidence that public contributions crowd out private contributions at a rate significantly less than dollar-for-dollar. Accordingly, even low levels of public contributions can significantly raise total funding. There is strong evidence that people give more when institutions activate psychological mechanisms, such as concerns about reciprocity and fairness, that play central roles in giving. From a policy perspective, this suggests that relatively inexpensive strategies involving advertising and the promotion of community leadership may deserve greater emphasis.

5.E. Looking Ahead

Given the crucial role that public goods and externalities play in many important policy problems, one of the main challenges ahead for public economics is to build and test better models of public goods, and to apply them to basic questions in public finance, political economy, and mechanism design. We are still far from a satisfactory model of public goods that can become a new workhorse for economic applications across the board. However, based on the rapidly growing body of evidence concerning the psychological and neural processes at work, including research on the neural basis of empathy, punishment, and cooperation (see DeQuervain et. al. [2004], McCabe et. al. [2001], Singer et. al. [2004], and Rilling [2002,2004]), we are optimistic that such a framework is on the horizon. Given the number of likely forces at work – reciprocity, social norms, social emotions, social signaling, and so forth – it seems likely that a relatively complex and multifaceted approach is needed. Yet it is also likely that the discovery of new organizing principles will permit useful simplifications that render the problem more tractable.

In focusing here on individual behavior, we have largely neglected the role of philanthropic organizations. As Andreoni [2004] convincingly argues, it is also essential to understand the behavior of the philanthropic sector. A growing body of evidence shows that charities significantly stimulate giving, and that their activities respond both to government policy and to the behavior of other not-for-profit institutions. For example, Andreoni and Payne [2003] show that government grants to charities reduce fund-raising activities; Andreoni and Petrie [2004] document the role of charities in disseminating information; and Harbaugh [1998] provides evidence that charities exploit social signaling in their fundraising campaigns. For reviews of the economics of philanthropy, see Andreoni [2004] and Rose-Ackerman [1996].

6. The Road Ahead

In our view, Behavioral Public Economics has enormous potential, and has already demonstrated its value by making important contributions to critical policy discussions. We have emphasized that the behavioral perspective does not preclude coherent normative analysis. Indeed, in many cases, it is possible to modify and extend the tools of empirical welfare analysis without abandoning familiar methodological principles. We have also reviewed recent behavioral work concerning policies affecting saving, addiction, and public goods. Each of these literatures offers novel and important insights, as well as the potential for groundbreaking innovation.

The goal of this final section is to briefly highlight some critical directions for future research.

Better models. While the current generation of behavioral models improves the explanatory power of economic theory, many behavioral patterns remain unexplained. Among other things, recent research has deemphasized considerations for which satisfactory and tractable formal models are not yet available, such as framing effects, the adoption of rules-of-thumb, and other responses to environmental complexity. To study the policy implications of these phenomena, better models are required. For example, we need theories that explain how people adopt rules of thumb, and how they adapt these rules to new environments.

New types and sources of data. With sufficiently restrictive structural assumptions, it is possible to estimate positive and normative behavioral models using data only on choices. The use of non-choice data would potentially allow economists to estimate these models more reliably, and to formulate more discriminating and robust tests of competing alternatives.

Future research should examine the possibility of measuring preferences directly, instead of inferring them from choices. Self-reporting is a natural source of information about tastes. In

practice, there are several problems with self-reported preferences. First, when choices are not involved, questions about preferences are inherently hypothetical. There is some reason to believe that people do not give reliable answers to hypothetical questions (see, e.g., List [2001] and the references therein). For example, unless there is something at stake, they may not take these questions seriously. Second, true preferences may conflict with social and moral norms, leading subjects to either rationalize or report false preferences. Third, people may make mistakes in assessing their own preferences. For example, a sizable body of literature has documented systematic errors in affective forecasting (see Loewenstein and Schkade [1999] for a review). Fourth, context may affect an individual's ability to cognitively access his true preferences.

Despite considerable evidence that self-reporting is susceptible to these problems (see Schwartz and Strack [1999] for a review), there is cause for optimism. For the most part, the object of this body of work has been to identify experimental manipulations that lead to nonsensical self-reports. While this demonstrates that there are important pitfalls associated with the direct elicitation of preferences, it does not prove that this approach is worthless. As far as we know, there has been no systematic attempt to design elicitation protocols that are stable and resistant to manipulation.

Many other types of data merit consideration. Even without eliciting complete preferences, one can potentially learn whether an individual regards a particular choice as a mistake, whether his choices correspond to his intentions, or whether he systematically fails to follow through on plans (see, e.g., Choi et. al. [2004] and Bernheim [1995]). One can also elicit information about expectations and make comparisons with realizations (see, e.g., Bernheim [1988, 1989], Hurd and McGarry [2002], Loewenstein, O'Donoghue, and Rabin [2003]). Obviously, information along these lines raises many of the same concerns as self-reported preferences. Finally, economists have only just started to tap data on physical states, brain activity, and the like. While the value of neuroeconomic data remains largely unproven, the potential payoffs are high, and the possibilities are worth exploring.

Difficult issues in welfare economics. The nascent field of "Behavioral Welfare Economics" is far from settled. Many thorny issues remain. The following hypothetical problem illustrates one challenging issue. An individual is presented with a choice between two options, *A* and *B*. He is indifferent between them. However, his preferences change as a result of his choice. If he chooses *A*, he prefers *B* (call this the "*A* self"). If he chooses *B*, he prefers *A* (call this the "*B* self"). Suppose he chooses *A*. Since only the initial self and the *A* self actually exist, it seems natural to place no weight on the preferences of the *B* self. But if we place any weight

on the *A* self, *B* is the welfare optimum. Of course, if we enforce this choice through public policy, the *A* self vanishes and the *B* self materializes, in which case *A* is the welfare optimum. Is there a coherent way to resolve this ambiguity? See Bernheim and Rangel [2005a] for a more systematic treatment of welfare economics in behavioral settings.

Non-standard policies. In the standard model, public policy affects behavior only through its effect on information and budget constraints. A growing body of literature, partially reviewed in the previous three sections, suggests that policy can also have powerful effects on behavior through other channels. For example, it can provide or suppress cues, and it can alter the way decision problems are framed. If economists can develop reliable formal models of these effects, it should be possible to study the optimal design of unconventional economic policies (e.g., restrictions on advertising, warning labels, and clever manipulations of framing effects as in Thaler and Shefrin's [2004] Saving for Tomorrow Savings Plan™) with the same rigor as traditional tax and expenditure policy.

New applications. The interesting collection of papers in this volume show that, as time passes, economists are applying behavioral economics to increasingly wide range of economic problems. No doubt this trend will continue within the field of Public Economics. Many of the tools described in this paper should prove useful in understanding issues pertaining to poverty, crime, corruption, and other important topics.

References

Abrams, Burton A. and Mark A. Schmitz (1978) "The Crowding Out Effect of Government Transfers on Private Charitable Contributions," Public Choice, 29-39

Abrams, Burton A. and Mark A. Schmitz (1978) "The Crowding Out Effect of Government Transfers on Private Charitable Contributions: Cross Sectional Evidence," National Tax Journal, 563-68

Ainslie, George, and Varda Haendel (1983), "The Motives of the Will," in E. Gotteheil et. al. (eds.), Etiologic Aspects of Alcohol and Drug Abuse. Springfiled, Illinois: Thomas.

Akerlof, George (1991) "Procrastination and Obedience," American Economic Review, LXXXI, 1-19

Ameriks, John, Andrew Caplin, and John Leahy (2003) "Wealth Accumulation and the Propensity to Plan," Quarterly Journal of Economics, 1007-47

Andreoni, James (1988) "Privately Provided Public Goods in a Large Economy: The Limits of Altruism," Journal of Public Economics, 57-73

Andreoni, James (1989) "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," Journal of Political Economy, 1147-58

Andreoni, James (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm Glow Giving," Economic Journal, 464-77

Andreoni, James (1993) "An Experimental Test of the Public-Goods Crowding-Out Hypothesis," American Economic Review, 1317-27

Andreoni, James (1995) "Warm-glow versus Cold-pickle: The Effects of Positive and Negative Framing on Cooperation Experiments," Quarterly Journal of Economics, 2-21

Andreoni, James (1998) "Towards a Theory of Charitable Fundraising," Journal of Political Economy, 1186-1213

Andreoni, James (2004) "Philanthropy," forthcoming in the Handbook of Giving, Reciprocity, and Altruism," L.A. Gerard-Vared, Serge-Christopher Kolm and Jean Mercier Ythier eds, Elseveir/North-Holland.

Andreoni, James and Abigail Payne (2003) "Do Government Grants to Private Charities Crowd Out Giving or Fund-Raising?" American Economic Review, 93:792-812.

Andreoni, James and Ragan Petrie (2004) "Public Goods Experiments Without Confidentiality: A Glimpse Into Fund-raising," Journal of Public Economics, 88, 1605-23

Angeletos George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman and Stephen Weinberg (2001) "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," Journal of Economic Perspectives, 15(3):47-68

Ariely, Dan and Ken Wertenbroch (2002) "Procrastination, Deadlines and Performance: Using Precommitment to Regulate One's Behavior," 13(3):219-24

Banks, James, Richard Blundell, and Sara Tanner (1988), "Is there a Retirement-Savings Puzzle?" American Economic Review, 88(4):769-88.

Bayer, Patrick J., B. Douglas Bernheim, and J. Karl Scholz (1996) "The Effects of Financial Education in the Workplace: Evidence from a Survey of Employers," mimeo, Stanford University.

Becker, Gary (1974) "A Theory of Social Interactions," Journal of Political Economy, 82:1063-93

Becker, Gary and Kevin Murphy (1988) "A Theory of Rational Addiction," Journal of Political Economy, 96(4):675-700

Benabou, Roland and Marek Pycia [2002] "Dynamic Inconsistency and Self-Control," Economics Letters, 77:419-24

Bergstrom, Theodore, Lawrence Blume, and Hal Varian (1986) "On the Private Provision of Public Goods," Journal of Public Economics, 25-49

Bernartzi, Sholomo (2001) "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," Journal of Finance, 56(5):1747-64

Bernartzi, Shlomo and Richard Thaler (2001) "Naïve Diversification Strategies in Defined Contribution Savings Plans," American Economic Review, 91(1):71-98

Bernheim, B. Douglas (1986) "On the Voluntary and Involuntary Provision of Public Goods," American Economic Review, 789-93

Bernheim, B. Douglas (1988), "Social Security Benefits: An Empirical Study of Expectations and Realizations," in E. Lazear and R. Ricardo-Campbell (eds.), Issues in Contemporary Retirement. Hoover Instution Press: Stanford, 312-345.

Bernehime, B. Douglas (1989), "The Timing of Retirement: A Comparison of Expectations and Realizations," in D. Wise (ed.), The Economics of Aging. NBER-University of Chicago Press: Chicago, 335-355.

Bernheim, B. Douglas (1994) "Personal Saving, Information, and Economic Literacy: New Directions for Public Policy," in *Tax Policy for Economic Growth in the 1990s*, Washington, D.C.: American Council for Capital Formation

Bernheim, B. Douglas (1995) "Do Households Appreciate Their Financial Vulnerabilities? An Analysis of Actions, Perceptions, and Public Policy," in *Tax Policy for Economic Growth in the 1990s*, Washington, D.C.: American Council for Capital Formation, 1-30.

Bernheim, B. Douglas (1998) "Financial Illiteracy, Education, and Retirement Saving," in *Living with Defined Contribution Pensions*, Olivia S. Mitchell and Sylvester J. Schieber, editors, Philadelphia: University of Pennsylvania Press, 38-68

Bernheim, B. Douglas, Lorenzo Forni, Jagadeesh Gokhale, and Laurence Kotlikoff (2002), "An Economic Approach to Setting Retirement Saving Goals," in Olivia Mitchell, Zvi Bodie, Brett Hammond, and Steve Zeldes (eds.), Innovations in Financing Retirement. Philadelphia: University of Pennsylvania Press, 77-105.

Bernheim, B. Douglas, and Daniel M. Garrett (2003), "The Effects of Financial Education in the Workplace: Evidence from a Survey of Households," Journal of Public Economics, 87(7-8):1487-1519.

Bernheim, B. Douglas, Daniel M. Garrett, and Dean Maki (2001), "Education and Saving: The Long-Term Effects of High School Financial Curriculum Mandates," Journal of Public Economics, 80(3):435-465.

Bernheim, B. Douglas and Antonio Rangel (2004) "Addiction and Cue-Triggered Decision Processes," American Economic Review, 94(5):1558-90

Bernheim, B. Douglas and Antonio Rangel (2005a) "Behavioral Welfare Economics," manuscript

Bernheim, B. Douglas and Antonio Rangel (2005b) "Savings and Cue-Triggered Decision Processes," manuscript

Bernheim, B. Douglas, and Antonio Rangel (2005c), "From Neuroscience to Public Policy: A New Economic View of Addiction," mimeo, Stanford University.

Bernheim, B. Douglas, Debraj Ray, and Sevin Yeltekin (1999), "Self-Control, Saving, and the Low Asset Trap," mimeo, Stanford University.

Bernheim, B. Douglas, Jonathan Skinner, and Steven Weinberg (2001), "What Accounts for the Variation in Retirement Wealth Among U.S. Households?" American Economic Review, 91(4), 832-857.

Berridge, K. (1996) "Food Reward: Brain Substrates of Wanting and Liking," Neuroscience and Biobehavioral Reviews, 20(1):1-25

Berridge, K. (1999) "Pleasure, Pain, Desire, and Dread: Hidden Core Processes of Emotion," in *Well-Being: The Foundations of Hedonic Psychology*, D. Kaheman, E. Diener, and N. Schwarz, editors, 525-57

Berridge, Kent, and Terry Robinson (1998) "What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?", Brain Research Review, 28:309-69

Berridge, Kent and Terry Robinson (2003) "Parsing Reward," Trends in Neuroscience, 26(9):507-513

Beshears, John, James Choi, David Laibson, and Brigitte C. Madrian (2005) "Early Decisions: A Regulatory Framework," forthcoming in Swedish Policy Review

Bhattacharya, Jay, and Darius Lackdawalla (2004), "Time-Inconsistency and Welfare," mimeo, Stanford University.

Blinder, Alan S. (1974). Toward an Economic Theory of Income Distribution. Cambridge, MA: MIT Press.

Bolton, Gary E., and Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition," American Economic Review, 20, 166-93.

Camerer, Colin (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Russell Sage Foundation.

Center for Disease Control (1993) "Smoking-Attributable Mortality and Years of Potential Life Lost - United States, 1990," Morbidity and Mortality Weekly Report, 42(33):645-8.

Chaloupka, F. and K. Warner (2001) "The Economics of Smoking", in Handbook of Health Economics, J. Newhouse and D. Cutler, editors, North-Holland

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2004a) "Optimal Defaults and Active Decisions," manuscript

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2004b) "Saving For Retirement on the Path of Least Resistance," manuscript

Clotfelter, Charles (1985). *Federal Tax Policy and Charitable Giving*. Chicago: University of Chicago Press.

Cohen, G. A. (1989), "On the Currency of Egalitarianism Justice," Ethics, 99(4):906-44.

Consumer Federation of America and the American Express Company (1991), *High School Competency Test Report of Findings*, Washington D.C.

Cornes, Richard and Todd Sandler (1999) "Easy Riders, Joint Production and Public Goods," Economic Journal, 94:580-98

Cronqvist, Henrik and Richard Thaler (2004) "Design Choices in Privatized Social Security Systems: Learning from the Swedish Experience," American Economic Review Papers and Proceedings, 94(2):424-28

DeQuervain, Dominique, et. al. (2004) "The Neural Basis of Altruistic Punishment," Science, 305:1254-58

Diamond, Peter (2005) "Optimal Tax Treatment of Private Contributions for Public Goods with and without warm-glow preferences," Journal of Public Economics, forthcoming

Diamond, Peter and Jerry Hausman (1994) " Contingent Valuation: Is Some Number Better than No Number?" Journal of Economic Perspectives, 8(4):45-64

Diamond, Peter and Botond Koszegi (2003) "Quasi-Hyperbolic Discounting and Retirement," Journal of Public Economics, 87(9-10):1839-72

Dockner, E. J., and G. Feichtinger (1993), "Cyclical Consumption Patterns and Rational Addiction," American Economic Review, 83:256-63.

Doyle, Robert J., Jr., and Eric T. Johnson (1991) *Readings in Wealth Accumulation Planning*, fourth edition, Bryn Mawr, Pennsylvania: The American College

Duflo, Esther and Emmanuel Saez (2002), "Participation and Investment Decisions in a Retirement Plan: the Influence of Colleagues' Choices," Journal of Public Economics, 85:121-48.

Duflo, Esther and Emmanuel Saez (2003) "The Role of Information and Social Interactions in Retirement Plans Decisions: Evidence from a Randomized Experiment," Quarterly Journal of Economics, 118(3): 815-842.

Engen, Eric M., William G. Gale, and John Karl Scholz (1994) "Do Saving Incentives Work?" *Brookings Papers on Economic Activity*, 85-151

Fang, Hanming and Dan Silverman (2002) "Time-Inconsistency and Welfare Program Participation: Evidence from the NLSY," manuscript

Farkas, Steve and Jean Johnson (1997) "Miles to Go: A Status Report on Americans' Plans for Retirement," Public Agenda

Feenberg, Daniel R. and Jonathan Skinner (1989) "Sources of IRA Saving," *Tax Policy and the Economy*, 25-46

Fehr, Ernst and Simon Gachter (2000) "Cooperation and Punishment in Public Good Experiments," American Economic Review, 980-94.

Fehr, Ernst and Simon Gachter (2002) "Altruistic Punishment in Humans," Nautre, 415:137-40.

Fehr, Ernst and Urs Fischbacher (2003) "The Nature of Human Altruism," Nature, 425, 785-9

Fehr, Ernst and Urs Fischbacher (2004) "Social norms and human cooperation," TRENDS in Cognitive Sciences, 8(4):185-90

Fehr, Ernst, and Klaus M. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation," Quarterly Journal of Economics, 114, 817-68.

Feldstein, Martin (1985) "The Optimal Level of Social Security Benefits," Quarterly Journal of Economics, 100(2):303-20

Fernandez-Villaverde, Jesus, and Arijit Mukherji (2002), "Can We Really Observe Hyperbolic Discounting?" mimeo, University of Pennsylvania.

Fisher, Irving, *The Theory of Interest*, London: MacMillan, 1930

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002) "Time Discounting and Time Preference: A Critical Review" Journal of Economic Literature, 40(2):351-401

Fudenberg, Drew and David K. Levine (2005) "A Dual Self Model of Impulse Control," manuscript

Gardner, Eliot and James David (1999) "The Neurobiology of Chemical Addiction,"in Getting Hooked: Rationality and Addiction, Jon Elster and Ole-Jorgen Skog, editors, Cambridge: Cambridge University

Goldstein, A. (2001) *Addiction: From Biology to Drug Policy, Second Edition*, New York: Oxford University Press

Goldstein, A. and H. Kalant (1990) "Drug Policy: Striking the Right Balance," Science 249:1513-21

Gross, David and Nicholas Souleles (2002) "Do Liquidity Constraints and Interest Rates Matter for Consumer Behaviors? Evidence from Credit Card Data, Quarterly Journal of Economics, 117(1):149-86

Gruber, Jonathan and Botond Koszegi (2001) "Is Addiction "Rational"? Theory and Evidence," Quarterly Journal of Economics, 116(4): 1261-1303

Gruber, Jonathan, and Botond Koszegi (2004), "A Theory of Government Regulation of Addictive Bads: Tax Levels and Tax Incidence for Cigarette Excise Taxation," Journal of Public Economics, 88(9-10):1959-1987.

Gul, Faruk and Wolfgang Pesendorfer (2001), "Temptation and Self-Control", Econometrica 69(6): 1403-35

Gul, Faruk and Wolfgang Pesendorfer (2004a) "Self Control, Revealed Preference and Consumption Choice," forthcoming in Review of Economic Dynamics

Gul, Faruk and Wolfgang Pesendorfer (2004b) "Self Control and the Theory of Consumption," forthcoming in Econometrica

Gul, Faruk, and Wolfgang Psendorfer (2005) "Harmful Addiction," forthcoming in Review of Economic Studies.

Hammermesh, Daniel S. (1984), "Consumption During Retirement: The Missing Link in the Life Cycle," Review of Economics and Statistics, 66(1): 1-7.

Harbaugh, William (1998) "What do donations buy? A model of philanthropy based on prestige and warm-glow," Journal of Public Economics, 67:269-84.

Harris, J.E. (1993) *Deadly Choices: Coping with Health Risks in Everyday Life*, New York: Basic Books

Hausman, Jerry A., and Lynn Paquette (1987), "Involuntary Early Retirement and Consumption," in Gary Burtless, ed., Work, Health, and Income Among the Elderly. Washington, DC: Brookings Institution Press, 151-75.

Holden, Sarah, Jack VanDerHei, and Carol Quick (2001) "401(k) Plan Asset Allocation, Account Balances, and Loan Activity in 1998," Investment Company Institute Perspective 6(1):

Hser, Y.I., D. Anglin, and K. Powers (1993) "A 24-year follow-up study of California narcotics addicts," Archives of General Psychiatry, 50:577-84

Hser, Y.I., V. Hollman, C. Grella, and M.D. Anglin (2001) "A 33 year follow-up of narcotics addicts," Archives of General Psychiatry, 58:503-8

Hurd, Michael, and Kathleen McGarry (2002), "The Predictive Validity of Subjective Probabilities of Survival," Economic Journal, 112:966-985.

Hurd, Michael, and Susanne Rohwedder (2003), "The Retirement-Consumption Puzzle: Anticipated and Actual Declines in Spending at Retirement," NBER Working Paper No. 9586.

Hyman, Steven and Robert Malenka (2001) "Addiction and the Brain: The Neurobiology of Compulsion and Its Persistence," Nature Reviews Neuroscience, 2:695-703

Iannacone, Laurence R. (1986), "Addiction and Satiation," Economic Letters, 21(1):95-99.

Imrohorglu, Selo, Ayse Imrohoroglu, and Douglas Joines (2003) "Time Inconsistent Preferences and Social Security," Quarterly Journal of Economics, 118(2):745-84

Isaac, Mark and James Walker (1988) "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," Quarterly Journal of Economics, 179-99.

Jones, Stephen R. G. (1984) *The Economics of Conformism*, Oxford: Basil Blackwell

Katona, George (1975). Psychological Economics. Amsertdam: Elsevier.

Kingma, Robert (1989) "An Accurate Measurement of the Crowd-out Effect, Income Effect, and Price Effect for Charitable Contributions," Journal of Political Economy, 1197-1207

Koszegi (2002) "Note: Any Model Features an Untestable Assumption About Preferences," Berkeley manuscript

Krusell, Per, and Anthony Smith (2003), "Consumption-Saving Decisions with Quasi-geometric Discounting," Econometrica, 71:365-75.

Krusell, Per, Burhanettin Kuruscu, and Anthony Smith (2000) "Tax Policy with Quasi-geometric discounting," International Economic Journal, 14(3):1-40

Krusell, Per, Burhanettin Kuruscu, and Anthony Smith (2002) "Equilibrium Welfare and Government Policy with Quasi-geometric Discounting," Journal of Economic Theory, 105:42-72

Krusell, Per, Burhanettin Kuruscu, and Anthony Smith (2001) "Temptation and Taxation," mimeo, Princeton University.

Laibson, David (1994), "Self-Control and Saving," mimeo, MIT.

Laibson, David (1997) "Golden Eggs and Hyperbolic Discounting," Quaterly Journal of Economics, 112(2):443-477

Laibson, David (1998), "Life-Cycle Consumption and Hyperbolic Discount Functions," European Economic Review, 42(3-5): 861-871.

Laibson, David (2001) "A Cue-Theory of Consumption," Quarterly Journal of Economics, 116(1): 81-120

Laibson, David, Andrea Repetto, and Jeremy Tobacman (2003) "A Debt Puzzle," in Phillip Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, eds., *Knowledge, Information, and Expectations in Modern Economics: In Honor of Edmund Phelps*, Princeton: Princeton University Press

Laibson, David, Andrea Repetto, and Jeremy Tobacman (2004) "Estimating Discount Functions from Lifecycle Consumption Choices," manuscript

Lazear, Edward P. (1994), "Some Thoughts on Saving," in David A. Wise, ed., Studies in the Economics of Aging. Chicago: University of Chicago Press and NBER, 143-69.

Ledyard, John (1995) "Public Goods: A Survey of Experimental Research," in Handbook of Experimental Economics, John Kagel and Alvin Roth eds, Princeton University Press, 111-94

List, John A., (2001), "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards," American Economic Review, 91(5):1498-507.

Loewenstein, George (1996) "Out of Control: Visceral Influences on Behavior," Organizational Behavior and Human Decision Processes, 65:272-92

Loewenstein, George (1999) "A Visceral Account of Addiction," in Rationality and Addiction, Jon Elster and Ole-Jorgen Skog, editors, Cambridge: Cambridge University Press

Loewenstein, George and Drazen Prelec (1998) "The Red and the Black: Mental accounting of savings and debt," Marketing Science, 1998, 17: 4-28

Loewenstein, George and David Schkade (1999) "Wouldn't It Be Nice? Predicting Future Feelings," in *Well-Being: the Foundations of Hedonic Psychology*, Daniel Kahneman, Ed Diener and Norbert Schwarz editors, New York, New York: Russell Sage Foundation, 85-104

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2003) "Projection Bias in Predicting Future Utility" , Quarterly Journal of Economics, 118(4):1209-1248

Loewenstein, George, Daniel Read, and Roy Baumister (eds) (2003), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, New York: Russell Sage Foundation

Loewenstein, George and Ted O'Donoghue (2004) "Animal Spirits: Affective and Deliberative Processes in Economic Behavior," manuscript

Long, James E. (1990) "Marginal Tax Rates and IRA Contributions," *National Tax Journal* 43(2):143-53

Lusardi, Annamaria (2000), "Saving for Retirement: The Importance of Planning," TIAA-CREF Institute, Issue No. 66.

Lusardi, Annamaria (2003), "Planning and Savings for Retirement," mimeo, Dartmouth College.

MacCoun, R. and P. Reuter (2001) *Drug War Heresies: Learning from Other Vices, Times, and Places*, Cambridge: Cambridge University Press

Madrian, Brigitte and Dennis Shea (2001) "The Power of Suggestion: Inertian in 401(k) Participation and Savings Behavior," Quarterly Journal of Economics, 116(4):1149-87

Mariger, Randall P. (1987), "A Life-Cycle Consumption Model with Liquidity Constraints: Theory and Empirical Results," Econometrica, May 1987, 55(3):533-57.

Masclet, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval (2003) "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanims," American Economic Review, 93(1):366-380

McCabe, Kevin, et. al. (2001) "A functional imaging study of cooperation in two-person reciprocal exchange," Proceedings of the National Academy of Sciences, 98(20):11832-35

McCaffrey, Edward and Joel Slemrod (2005) "Toward an Agenda for Behavioral Public Finance," manuscript

McClure, Sam et. al. (2004) "Separate Neural Systems Value Immediate and Delayed Monetary Rewards," Science 306: 503-507

Metcalfe, Janet, and Walter Mischel (1999) "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower," Psychological Review, 106(1):3-19

Mischel, Walter (1974) "Processes in delay of gratification," In Advances in Experimental Social Psychology, vol. 7, D. Berkowitz editor, 249-72

Mischel, Walter and B. Moore (1973) "Cognitive Appraisals and Transformations in Delay of Gratification," Journal of Personality and Social Psychology, 28:172-9

Mischel, W., Y. Shoda, and M. Rodriguez (1992) "Delay of Gratification in Children," in Choice Over Time, G. Loewenstein and J. Elster editors, New York: Rusell Sage

Modigliani, Franco, and Richard Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in *Post Keynsian Economics*, K.K. Kurihara, editor, New Brunswick, NJ: Rutgers University Press, 1954

National Institute on Alcohol Abuse and Alcoholism (2001) "Economic Perspectives in Alcoholism Research," Alcohol Alert, National Institutes of Health, No. 51

National Institute on Drug Abuse (1998) *The Economic Costs of Alcohol and Drug Abuse in the United States, 1992*, Bethesda, MD: National Institutes of Health

Nestler, E.J. (2001) "Molecular Basis of Long-term Plasticity Underlying Addiction," Nature Reviews Neuroscience, 2:119-28

Nestler, E. and RobertMalenka (2004) "The Addicted Brain," Scientific American, March:78-85

Norstrom, Thor, and Ole-Jorgen Skog (2005) "Saturday Opening and Alcohol Retail Shops in Sweden: An Experiment in Two Phases," Addiction, 100:767-776

O'Brien, C. (1976) "Experimental analysis of conditioning factors in human narcotic addiction," Pharmacological Review, 25:533-43

O'Brien, C. (1997) "A Range of Research-Based Pharmacotherapies for Addiction," Science, 278:66-70

O'Donoghue, Ted and Matthew Rabin (1999a) "Procrastination in Preparing for Retirement," in Henry J. Aaron, ed., Behavioral Dimensions of Retirement Economics, Washington, D.C.: Brookings Institution Press; New York, NY, 125-56

O'Donoghue, Ted and Matthew Rabin (1999b) "Doing It Now or Later," American Economic Review, 89(1):103-24

O'Donoghue, Ted and Matthew Rabin (2001) "Choice and Procrastination," Quarterly Journal of Economics, 121-60

O'Donoghue, Ted and Matthew Rabin (2005) "Optimal Sin Taxes," manuscript

Office of National Drug Control Policy (2001a) "What American Users Spend on Illegal Drugs," Washington, DC: Executive Office of the President (Publication No. NCJ-192334)

Office of National Drug Control Policy (2001b) "The Economic Costs of Drug Abuse in the United States," 1992-1998, Washington, DC: Executive Office of the President (Publication No. NCJ-190636).

O'Neill, Barbara (1993), "Assessing America's Financial IQ: Realities, Consequences, and Potential for Change," mimeo, Rutgers Cooperative Extension.

Ornstein, Stanley I. , and Dominique M. Hanssens (1985) "Alcohol Control Laws and the Consumption of Distilled Spirits and Beer," Journal of Consumer Research, 12:200-212

Orphanides, Athanasios, and David Zervos (1995) "Rational Addiction with Learning and Regret," Journal of Political Economy, 103: 739-58

Orphanides, Athanasios, and David Zervos (1995), "Myopia and Addictive Behavior," Economic Journal, 108:75-91.

Palfrey, Thomas and Jeffrey Prisbrey (1997) "Anomalous Behavior in Public Goods Experiments: How Much and Why?," American Economic Review, 829-46.

Paserman, Daniele (2002) "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," manuscript

Phelps, Edmund and Robert Pollack (1968) "On Second-Best National Savings and Game Equilibrium Growth," Review of Economic Studies, XXXV, 185-99

Rabin, Matthew (2002) "A Perspective on Psychology and Economics," European Economic Review, 46(4-5):657-85

Rainwater, Lee (1970) *Behind Ghetto Walls: Black Families in a Federal Slum*, Chicago: Aldine

Redish, A. David (2004), "Addiction as a Computational Process Gone Awry," Science, 306:1944-1947.

Rilling, James, et. al. (2002) "A Neural Basis for Social Cooperation," Neuron, 35:395-405

Rilling, James, et. al. (2004) "Opposing Bold Responses to Recriprocated and Unreciprocated Altruism in Putative Reward Pathways," Neuroreport, 15(16): 2539-243.

Robb, A. Leslie, and John B. Burbidge (1989), "Consumption, Income, and Retirement," Canadian Journal of Economics, 22(3):522-42.

Robbins, T.W. and Everitt B.J. (1999) "Interaction of depaminergic system with mechanisms of associative learning and cognition: implications for drug abuse," Psychological Science, 10:199-202

Roberts, Russell (1984) "A Positive Model of Private Charity and Public Transfers," Journal of Political Economy, 136-8

Robins, L. (1994) "Vietnam Veterans' Rapid Recovery from Heroin Addiction: a Fluke or Normal Expectation," Addiction, 1041-54

Robins, L., D. Davis, and D. Goodwin (1974) "Drug Use by U.S. Army Enlisted Men in Vietnam: a Follow-up on their Return Home," American Journal of Epidemiology, 235-49

Robinson. T. and K. Berridge (1993) "The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction," Brain Research Reviews, 18(3):247-91

Robinson, T. and K. Berridge (2000) "The psychology and neurobiology of addiction:
an incentive sensitization view," Addiction, Suppl 2:91-117

Robinson, Terry and Kent Berridge (2003) "Addiction," Annual Reviews of Psychology, 54:25-53

Roemer, John E. (1998). Equality of Opportunity. Cambridge, MA: Harvard University Press.

Rose-Ackerman, Susan (1996) "Altruism, Nonprofits, and Economic Theory," Journal of Economic Literature, XXXIV,701-28

Scholz, J. Karl, Ananth Seshadri, and Surachai Khitatrakun (2004), "Are Americans Saving Adequately for Retirement?" mimeo, University of Wisconsin, Madison.

Schultz, W. (2000) "Multiple Reward Signals in the Brain," Nature Reviews Neuroscience, 1:199-207

Schultz, W., P. Dayan, and P.R. Montague (1997) "A neural substrate of prediction and reward," 275:1593-99

Schultz, W. (1998) "Predictive reward signal of dopamine neurons," Journal of Neurophysiology, 80:1-27

Schwarz, Norbert, and Fritz Strack (1999) "Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications," in *Well-Being: the Foundations of Hedonic Psychology*, Daniel Kahneman, Ed Diener and Norbert Schwarz editors, New York, New York: Russell Sage Foundation, 61-84

Scitovsky, Tibor (1976). The Joyless Economy. Oxford: Oxford University Press.

Sefton, Martin, Robert Shupp, and James Walker (2002) "The Effect of Rewards and Sanctions in the Provision of Public Goods," CEDEX Research Paper, 2002, University of Nottingham

Sen, Amartya (1992). Inequality Reexamined. Cambridge, MA: Harvard University Press.

Shang, Jeng and Rachel Croson (2005) "Field Experiments in Charitable Contributions: The Impact of Social Influence on the Voluntary Provision of Public Goods," manuscript

Shefrin, H.M. and Richard Thaler (1981) "An Economic Theory of Self-Control," Journal of Political Economy

Shiv, B. and A. Fedorikin (1999) "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making," Journal of Consumer Research, 26:72-89

Singer, Tania, et. al. (2004) "Brain Responses to the Acquired Moral Status of Faces," Neuron, 41(4):653-62.

Stack, Carol B. (1974) *All Our Kin: Strategies for Survival in a Black Community*, New York: Harper & Row

Steinberg, R. (1987) "Voluntary Donations and Public Expenditures in a Federalist System," American Economic Review, 77:24-36

Stigler, George and Gary Becker (1977) "De Gustibus Non Est Disputandum," American Economic Review, 67:76-90

Strotz, Robert H. (1956) "Myopia and Inconsistency in Dynamic Utility Maximization," Review of Economic Studies, XXIII, 165-80

Sugden, Robert (2004), "The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences," American Economic Review, 94(4):1014-1033.

Summers, Lawrence H. (1986) "Summers Replies to Galper and Byce on IRAs," *Tax Notes*, 31(10):1014-1016

Sunstein, C. and Thaler, R. (2003) "Libertarian Paternalism," The American Economic Review, Vol. 93:2, 175-179

Thaler, Richard (1985) "Mental Accounting and Consumer Choice," Marketing Science, ***

Thaler, Richard and Shlomo Benartzi (2004) "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," Journal of Political Economy, 112(1):S164-187

Thaler, Richard and H.M Shefrin (1981) "An Economic Theory of Self-Control," Journal of Political Economy, 89(2): 392-406

Tigerstedt, Christoffer, and Caroline Sutton (2000) "Exclusion and Inclusion – Saturday Closings and Self-Service Stores," in *Broken Spirits. Power and Ideas in Nordic Alcohol Control*, edited by Pekka Sulkunen, Caroline Sutton, Christoffer Tigerstedt and Katariina Warpenius, Nordic Council for Alchohol and Drug Research, 185-201.

Trosclair, A., C. Huston, L. Pederson, and I. Dillon (2002) "Cigarette Smoking Among Adults — United States, 2000," Morbidity and Mortality Weekly Report, 51(29): 642-645.

Tversky, Amos and Daniel Kahneman (1986) "Rational Choice and the Framing of Decisions," Journal of Business, 59(4):5251-78

United States Census Bureau (2001) *Statistical Abstract of the United States*, Washington, DC: US Government Printing Office

United States Department of Health and Human Services (1994) "Preventing Tobacco Use Among Young People: A Report of the Surgeon General," National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health

Venti, Steven F., and David A. Wise (1992) “Government Policy and Personal Retirement Saving,” *Tax Policy and the Economy* 6, 1-41

Walstad, William B. and Max Larsen (1992), "A National Survey of American Economic Literacy," National Council on Economics Education.

Walstad, William B. and John C. Soper (1988), "A Report Card on the Economic Literacy of U.S. High School Students," *American Economic Review* 78(2):251-56.

Warshawsky, Mark (1987), "Sensitivity to Market Incentives: The Case of Policy Loans," *The Review of Economics and Statistics* 69(2):286-95.

Warr, Peter (1982) “Pareto Optimal Redistribution and Private Charity,” Journal of Public Economics, 131-38

Wertenbroch, K. (1998) ‘Consumption and Self-Control for Rationing Purchase Quantities of Virtue and Vice,” Marketing Science, 17:317-37

Whyte, William F. (1943) *Street Corner Society*, Chicago: The University of Chicago Press

Wickelgren (1997) “Getting the Brain’s Attention,” Science, 278:35-37

Wise, Roy (1989) “The Brain and Reward,” in *The Neuropharmacological Basis of Reward*, J.M. Liebman and S.J. Cooper, editors, New York: Oxford University Press, pp. 377-424

Yerak, Becky (2001) “Program helps gamblers quit,” The Detroit News, Sunday, Dec 2.