Distinguishing Between Cognitive Biases: Beliefs vs. Time Discounting in Welfare Program Participation*

Hanming Fang
Yale University

Dan Silverman
University of Michigan

March 18, 2005

Abstract

As economists introduce more and more cognitive biases into economic models, an important question is whether we can distinguish these biases empirically. We present a simple model of welfare program participation that nests two well-documented cognitive biases: projection bias and present bias. We argue that agents with present bias and projection bias will exhibit different attitude toward time limits and other welfare eligibitlity restrictions, both before and after such restrictions are implemented. To the extent that such attitudes can be accurately elicited and measured, we argue that we can use attitudinal data to distinguish present bias and projection bias models.

*An earlier draft of the paper was presented in the conference "Behavioral Public Finance: Toward a New Agenda" at the University of Michigan on April 23-24, 2004. We are grateful to George Loewenstein, Edward McCaffery, Joel Slemrod and other conference participants for helpful comments. All remaining errors are our own.

1 Introduction

Economists have recently taken increased interest in a number of cognitive biases and heuristics first documented by psychologists.[1] In theoretical studies, economists typically introduce these biases and heuristics into stylized models with an aim of understanding how small, but psychologically relevant, deviations from the standard economic framework can influence decisions such as saving and consumption (Harris and Laibson 2001), investment (Barberis and Huang 2001), and labor supply (Fang and Silverman 2004a). In empirical studies, economists have followed two basic strategies. The first strategy is to derive distinctive empirical implications from a model of a particular bias or heuristic and then check if the data are qualitatively consistent with the bias model's predictions but inconsistent with the standard model's. (See, e.g., Babcock, et al., 1997, Genesove and Mayer, 2001 and Della Vigna and Malmandier, 2004.) The second strategy involves estimating structural models that allow a particular bias and attempt to measure the degree of that bias and its implications (Fang and Silverman, 2004b, Paserman, 2003). Unlike the first, this second strategy assumes an explicit model that permits simulations of the behavioral and welfare consequences of counterfactual policy experiments.

An important motivation for incorporating cognitive biases and heuristics in economic analyses derives from their implications for public policy. When a public economist evaluates a policy, the typical first step is to consult the formalization of Adam Smith's invisible hand in the "First Fundamental Theorem of Welfare." According to the theorem, when competitive markets exist and their participants share information commonly, the allocations of those markets are efficient; no other feasible allocation could make one person better off without making someone else worse off. It follows that if markets are missing, or imperfectly competitive, or if economic agents are acting with incomplete information, policy interventions may be justified on efficiency grounds.

The implications of cognitive biases and heuristics for decision-making form the basis of another rationale for policy interventions into economic activity. Biases and heuristics may drive a wedge between normative or long-term preferences and revealed preferences. That is, biases and heuristics can make what an individual actually chooses to do different from what she would do if she perceived

[1] The biases studied by economists include present-biased discounting (Strotz 1956, Laibson 1997, O'Donoghue and Rabin 1999), confirmatory bias (Rabin and Schrag, 1999), loss aversion (Genesove and Mayer 2001), mental accounting (Barberis and Huang 2001), and projection bias (Loewenstein, O'Donoghue and Rabin 2003).

her world without bias, or made her choices from a temporal distance, or could pay careful attention to her decisions at zero cost. If biases or heuristics lead economic agents to decisions that are at odds with their normative or long-term preferences, then public policy interventions could in principle make some individuals better off, while making no one worse off. Thus, even in cases where perfect markets exist, the influence of cognitive biases and heuristics may justify public policy interventions on efficiency grounds.

While investigations into the relevance of cognitive biases and heuristics for economic decision making may have profound policy implications, this research agenda faces a fundamental difficulty: given the large number of deviations from strict rationality that have been documented by psychologists, how should researchers determine which ones apply to which economic settings?

So far, both the theoretical and the empirical studies in economics have tended to investigate the implications of cognitive biases and heuristics *one bias at a time*.[2] The strategy of limiting attention to one potential bias may, in some cases, be justified by *a priori* indications of the primacy of that bias in determining behavior. Even when there is no *a priori* logic that favors analysis of one bias rather than another, there are clear advantages to adopting an incremental approach. For theory, isolating the influence of a minimal deviation from the standard framework has intrinsic interest and provides greater potential for clarity and tractability. An additional advantage of limiting empirical studies to one bias at a time derives from a concern for distinguishing among the effects of various biases. In some quite standard settings it is impossible to empirically distinguish between a model of even a single bias and a traditional model (see, e.g., Barro 1999). Distinguishing among multiple models of bias at once may present substantial challenges.

Nevertheless, in many cases several biases might plausibly explain behavior. In these cases, the obvious question for empirical research is whether we can distinguish these various biases from each other, and from a traditional model, using readily available data. If readily available data is not sufficient to distinguish among the biases, it then becomes important to investigate what additional data should be collected. Distinguishing among biases is important because, as we will demonstrate below, different biases may lead to very different policy recommendations.

In this paper, our primary task is to determine whether certain biases can be distinguished using data on labor supply and welfare program participation. In the context of a simple model of

[2] Barberis and Huang (2001) is an exception.

work and welfare program participation we investigate whether it is possible to distinguish between two psychological biases: time inconsistent discounting in the form of present-biased (hyperbolic) time preferences and non-rational beliefs in the form of projection bias. We show formally that indeed these two simple models of these biases can be distinguished from each other, and from a conventional model using standard data. In addition, we highlight a novel distinction between the two biases and argue that individuals under the two biases may exhibit different attitudes and attitudinal changes over time toward welfare eligibility restrictions such as time limits. To the extent that such data may be collected and analyzed, these differences in attitudes and changes in attitude provide an unexplored channel that researchers may exploit to distinguish biases in belief and time discounting.

The remainder of the paper is structured as follows. Section 2 uses three concrete examples from the behavioral economics literature to demonstrate both the potential importance of cognitive biases for public policy evaluation and the value of being able to distinguish empirically among biases. Section 3 describes several alternative methods for distinguishing among biases. Section 4 provides an informal discussion of the models we use to illustrate the effects of present-biased time discounting and belief biases on labor supply; their formal details are given in Appendix A. Section 5 summarizes our analysis of how the present bias and projection bias models may be distinguished either using standard data on choices and outcomes, or using data on (changes in) attitudes towards welfare policy reforms such as time limits. Formal treatments of these issues are provided in Appendices B and C. Finally, Section 6 concludes.

2 Cognitive Biases and Public Policy

As noted in the Introduction, the existence of cognitive biases and heuristics creates the potential for important efficiency gains from public policy interventions – even in settings where, if agents were strictly rational, the first fundamental theorem of welfare would hold. The behavioral economics literature on present-biased preferences provides a set of concrete examples of both the potential for these efficiency gains, and the potential importance of multiple biases.

2.1 Present-biased Preferences and Public Policy

The recent literature that studies the influence of present-biased time preferences on economic decision making is a prominent example of how economists have introduced cognitive biases into their analyses. By way of background, present-biased time preferences are a simple way of modeling the intuitive notion of a taste for immediate gratification and the resulting problems of self-control. Following the lead of Laibson (1997), who built on earlier work by Strotz (1956) and Pollack (1968), economists have used quasi-hyperbolic time discounting models to study a variety of economic questions.

An important feature of present-biased time preferences is that they provide a simple way for economists to explain choice reversals commonly observed in experimental and survey research: subjects choose the larger and later of two prizes when both are distant in time, but prefer the smaller and earlier one as both prizes draw nearer to the present.[3] The essence of such choice reversals is conveyed by two simple examples. In the first example, imagine that subjects are asked on February 1 to choose between spending seven hours on the tax return (an unpleasant task) on April 1 versus spending eight hours on April 15. Almost everyone would prefer to spend seven hours on April 1; but when April 1 arrives, most subjects, facing the same two alternatives, would put the work off until till April 15. That is, individuals have a tendency to procrastinate on unpleasant tasks. In the second example, imagine that subjects are asked to choose whether to receive a prize of a $100 certified check available immediately or to have a $200 certified check that could not be cashed for 2 years. A majority of the subjects choose to receive the $100 certified check that can be cashed immediately. However, the same people do not prefer a $100 certified check that could be cashed in 6 years to a $200 certified check that could be cashed in 8 years (see Anslie and Haslam 1992). That is, when faced with a pleasant reward, individuals show a short-run desire for instantaneous gratification.

An important implication of present-biased preferences is that revealed rates of time discount tend to decline with time and thus introduce the potential for both problems of self control and

[3] See Ainslie (1992) and several papers in Loewenstein and Elster (1992) for reviews of these experiments. Frederick, Loewenstein and O'Donoghue (2002) provide a comprehensive critical survey of time discounting and time preferences. Rubinstein (2003) provides alternative explanations of the choice reversal phenomenon, and presents a set of experiments contradicting the implications from hyperbolic discounting formulation.

utility gains from restricting choice sets. These gains from fewer choices are in stark contrast to the implications of the time-consistent, exponential time discounting that economists conventionally assume; with time-consistent discounting, the restriction of an individual's choice set can never make her better off in a single-agent decision problem.

2.1.1 Examples of Efficiency Gains from Policy

Present-biased preferences have been used to explain a variety of otherwise anomalous economic behaviors. Of primary importance for this article, however, are the implications that this bias has for public policy. We examine these implications in the context of three examples from the behavioral economics literature.

Example 1: Gruber and Koszegi (2001) and the Design of Cigarette Taxes

In a recent paper, Jonathan Gruber and Botond Koszegi consider the influence of present-biased preferences on cigarette smoking decisions and optimal tax policy. Their paper combines a theoretical investigation with a quantitative evaluation of the impact of intentional smoking choices on the well-being of smokers. Cigarette taxation is, of course, an example in the standard public economics literature of the taxation of activities with externalities. By definition, the performance of this type of activity by one individual has spillover effects on others, and those spillovers are not reflected in the price of the activity. When the spillovers have positive value, the activity is overpriced and it will be under-consumed from a social standpoint; theory argues it should be subsidized. If the spillover effects are negative, the activity is underpriced and will be over-consumed, and should thus be taxed.

The standard analyses of smoking externalities focus on the health costs of smoking that are not included in the price of cigarettes, but are borne by both smokers and non-smokers. Gruber and Koszegi's startling finding is that, if consumers have present-biased preferences consistent with experimental psychology, the costs of smoking's standard externalities may be dwarfed by what the authors call "aggregate internalities;" that is, the health costs *to the smoker* from over-smoking due to self control problems. The key point is that present-biased preferences may harm smokers by causing them to over-smoke with respect to their own normative or long-term preferences where, again, these are the preferences smokers would reveal if they saw their world without bias and could make decisions from a temporal distance. The policy implications of such internalities are profound.

Gruber and Koszegi's calculations suggest that an optimal tax response to smoking internalities would indicate cigarette taxes that are many times larger than current levels.

Example 2: Choi *et al.* (2004) and the Design of 401(k) Saving Plans

The paper by Choi, *et al,* in this volume presents compelling evidence that, when making their decisions regarding contributions to 401(k) retirement savings accounts, individuals often follow "the path of least resistance," and end up saving less than they had planned to save. More specifically, the evidence collected by Choi, et al., indicates that individuals often choose to do "nothing" when it comes to retirement savings decisions and simply adopt the default options set by their employers. As a result, these default saving options (including both saving rates and portfolio choices) can have a profound influence on saving outcomes.

One interpretation of the tendency to do nothing about important life decisions is that it represents a form of procrastination. Self-control problems may lead individuals continually to postpone the research and thinking required to make an important saving choice. Again, time-inconsistent preferences can account for this counterproductive urge to procrastinate even, perhaps especially, on highly consequential choices.[4] Thus, in this saving context, present-biased preferences may again justify policy interventions that facilitate self regulation or, as in the case of setting positive default saving rates, exploit tendencies to procrastinate in order to achieve superior saving outcomes.[5]

Example 3: Fang and Silverman (2004a,b) and the Design of Welfare Policy.

In Fang and Silverman (2004a) we investigated the implications of present-biased preferences for the labor supply and welfare program participation decisions of single mothers with children. Our analysis was motivated by the common claim that long-term dependence on welfare is suboptimal, not just for taxpayers but also for the recipients themselves. That claim has, in turn, led some policy makers to argue that policies such as time limits, which restrict an individual's cumulative benefits to a certain number of years, and workfare, which provides benefits only to those who satisfy work requirements, may actually benefit welfare recipients. As suggested above, this claim runs

[4] O'Donoghue and Rabin (2001) show how consumers with time-inconsistent preferences may procrastinate more when pursuing important goals than unimportant ones.

[5] An example of one such successful policy is described in Benartzi and Thaler (2004).

counter to a fundamental property of single-agent economic decision problems with standard time preferences: the restriction of an individual's choices can, at best, leave her well-being unchanged; it can never make her better off.

Our paper showed that, when single mothers have present-biased time preferences, time limits could benefit the welfare-eligible by providing them with a commitment to work – a commitment that alleviates problems of self control. We explained why some, who according to their long-term preferences would have preferred to work (and eventually consume) more, would instead choose welfare. Thus, if preferences are present-biased, in some cases, welfare policy interventions could generate efficiency gains, even if markets were perfect. In fact, when we examined data on labor supply and welfare program participation among single women with children, and estimated a model that allowed but did not assume a simple form of present-biased time preferences (Fang and Silverman, 2004b), we found evidence of present-bias among single mothers with children. More specifically, our estimates of the time-discount function indicated significantly higher rate of time discount for the near versus the more distant future. We also found, however, that when we used the estimated model to simulate the effects of policy changes, the imposition of time limits and work requirements failed to make most single mothers better off despite the present-biased preferences. In this context, it appeared that while some single-mothers sometimes could benefit from an exogenously imposed commitment to work, time limits and workfare were too crude instruments to realize this benefit. Too often these restrictions forced women to work when their best option was instead welfare.

2.2 The Implications of Alternative Biases

Present-biased time preferences are just one of many deviations from a strictly rational model of decision making that have been documented by psychologists. In this subsection, we consider how the presence of other cognitive biases, perhaps in addition to present-biased time preferences, might influence behavior and therefore the evaluation of public policies relevant to the previous examples. We begin with a bit of background on two well-documented belief biases: optimism and projection bias.

2.2.1 Optimism Bias

There is considerable evidence that individuals consistently exhibit a bias towards optimism (overconfidence) in beliefs. Psychological evidence of overconfidence is first and foremost reflected in the *"above median effect,"* whereby well over half of survey respondents typically judge themselves in possession of more desirable attributes than fifty percent of other individuals in the relevant comparison group. In Svenson (1981), 81 American and 80 Swedish students were asked to judge their skill in driving and how safe they were as drivers. Svenson found that 92.8% of American and 68.7% of Swedish subjects rated themselves as safer than 50% of other drivers. In Larwood and Whittaker (1977), 72 undergraduate management students and 48 presidents of New York state manufacturing firms are asked to rate themselves relative to their classmates or fellow presidents in IQ, likelihood of success, predicted growth in a hypothetical marketing problem, etc. The results indicate an astonishing level of overconfidence: of the 72 students, only 10 felt that they were merely of average intelligence relative to their own classmates and only 2 thought themselves below average; and only 18 of the 72 subjects predicted that their hypothetical firm's sales would be below the industry average. The executive sample also predicted inordinate success, even though more moderate than the students.

Psychological evidence of overconfidence is also reflected in the *"fundamental attribution error"* (Aronson 1994); that is, people tend to attribute their successes to ability and skill, but their failure to bad luck or to factors out of their control. Such self-serving biases are bound to reinforce overconfidence. Psychologists have gathered a great deal of evidence indicating that we take credit for the good and deny the bad. For example, students who do well on an exam attribute their performance to ability and effort, whereas those who do poorly attribute it to a poor exam or bad luck (Arkin and Maruyama 1979); gamblers perceive their success as based on skill and their failure as a fluke (Gilovich 1983); when married couples estimate how much of the housework each routinely did, their combined total of housework performed amounts to more than 100 percent - in other words, each person thinks he or she did more work than their partner think he or she did (Ross and Sicoly 1979); two-person teams performing a skilled task accept credit for the good scores but assign most of the blame for the poor scores to their partner (Johnston 1967); when asked to explain why someone else dislikes them, college students take little responsibility for themselves (i.e., there must be something wrong with this other person), but when told that someone else likes

them, the students attributed it to their own personality (Cunningham, Starr and Kanouse 1979).

2.2.2 Projection Bias

Another particularly relevant and well-documented bias is the deviation from rational beliefs known as projection bias. Projection bias refers to a tendency to mispredict future utilities. Specifically, experimental evidence indicates that, in many contexts, individuals understand qualitatively how their tastes will change with time or circumstances, but systematically underestimate the magnitudes of these changes. In particular, individuals tend to exaggerate the degree to which their future tastes will resemble their current tastes.

An example of a projection bias experiment illuminates the nature of the bias. In Read and van Leeuwen (1998), 200 workers were asked to choose between a healthy snack (fruit) and an unhealthy snack (candy, nuts, chips) that they would receive, free of charge, a week later.[6] The workers were told either that their snack would be delivered in the late afternoon, when they should be hungry, or just after lunch, when satiated. Some workers were asked to make this advance choice in the afternoon, when they were hungry, while the remainder was asked just after lunch. Table 1 describes the workers' choices depending on their current and anticipated future states of hunger.

		future state	
		Hungry	Satiated
current	Hungry	78%	56%
state	Satiated	42%	26%

Table 1: Percent of Subjects Making an Advance Choice of an Unhealthy Snack, from Read and van Leeuwen (1998).

Those who anticipated being hungry when the snacks were to be delivered were more likely to choose the unhealthy snack if they were currently hungry. The same is true of those who anticipated being satiated. In this way, the current state of hunger appears to influence predictions about future preferences. Those who are currently hungry, and who presumably have an greater taste for the unhealthy snack, appear to project their current tastes onto their future ones.[7] The respondents

[6] This experiment is cited in Loewenstein et al. (2003).

[7] A week later, the experimenters arrived at the promised times and allowed the subjects to make an immediate

are aware that their preferences are state-dependent. For example, the currently hungry choose the healthy snack more often when they anticipate being satiated. Judging by the choices of the currently satiated, however, they seem to seriously underestimate the change in their preferences.

2.2.3 Alternative Biases and Public Policy

Returning to our three examples from the behavioral economics literature, we now consider how optimism and projection biases might influence behavior and therefore public policy design.

Example 1: Gruber and Koszegi (2001) and the Design of Cigarette Taxes.

It seems natural that the smoking decision would be strongly influenced by time discounting. The rewards of smoking – the flavor, relief of cravings, feeling of relaxation – are immediate, while the health costs are often borne in a distant future. Thus a bias towards immediate gratification might profoundly influence behavior. However, optimism and projection biases might also influence the smoking decision in important ways, and lead to either to over- or under-smoking with respect to normative or long-run preferences.

Consider first optimism bias. If, knowing the distribution of health risks from smoking, the typical consumer thinks herself less likely than average to suffer these costs, then optimism bias would lead to over-smoking for a reason other than self-control problems. Specifically, if the consumer could view her environment without optimism bias, her preference would be to smoke less. If policy makers give priority to this unbiased preference, optimal cigarette tax designs that ignored the influence of optimism could, therefore, result in too much smoking.

Alternatively, projection bias might also importantly influence smoking decisions. Recall that projection bias leads consumers to exaggerate the extent to which their future tastes will resemble their current ones. Suppose that young people overestimate the degree to which their current tastes for vigorous physical activities and energy will apply at older ages; and suppose older people in fact place a lower value on physical vigor. To the extent that poor health curtails physical activities, the disutility of the future health costs of smoking may receive too much weight in the smoking decisions of young people, and they may smoke too little. Optimal cigarette tax policies that

choice of snack, regardless of the choice they had made in advance. Those who were currently hungry chose the unhealthy snack more often than those who were satiated. However, both the hungry and the satiated opted for the unhealthy snack much more often than they had when the same choice was offered a week in advance.

ignored this form of projection bias would result in overtaxation and socially suboptimal levels of smoking.

Example 2: Choi *et al.* (2004) and the Design of 401(k) Saving Plans.

As with the smoking decision, it is natural to think that saving choices would be influenced to a large degree by the nature of time discounting. Analogous to the effort required not to smoke, the costs of foregone consumption and planning for retirement are immediate, but the rewards are long delayed. Thus, in particular, a taste for immediate gratification could generate costly deviations from the choices that are optimal with respect to long-term preferences. Several studies have found, however, that variation in time discount rates explains relatively little of the variation in wealth. See, e.g., Barsky, et al. (1997), Bernheim, et al. (2001) and Ameriks, et al. (2003). Moreover, there is reason to think that some of the other well-documented cognitive biases that could influence saving decisions and therefore optimal public policy design.

For example, the optimism bias could, as in the case of smoking, lead to sub-optimal saving decisions. If, for instance, the average person believes her income will grow at an above average rate, she will tend to save too little for retirement. Certain forms of projection bias, on the other hand, would seem to generate too much saving. One scenario in which projection bias may produce over-saving is related to the finding that, according to a variety of estimates, the average household reduces its consumption by 10-30% at retirement.[8] If this decline reflects intentional behavior, and if working-age families with children overestimate the extent to which their tastes for consumption in retirement will resemble their current tastes, these families will tend to over-save with respect to their normative preferences.

The potential influence of these other cognitive biases on saving choices may prove important for policy design especially if, as seems natural, policy makers use surveys to elicit the saving preferences of individuals, and then use these stated preferences to determine default saving rates. If, indeed, time-inconsistent preferences are leading to procrastination on saving choices, then eliciting preferences about relatively distant saving choices and setting defaults accordingly could generate important utility gains. This strategy for setting defaults has important potential in this case because a distinguishing characteristic of time-inconsistent discounting is that normative preferences over choices are little affected if those choices occur in the relatively distant future.

[8] See, e.g., Banks, et al. (1998) and Bernheim (2001).

Other cognitive biases such as optimism and projection bias, do not have this feature. If these biases apply, even (and perhaps especially) stated preferences over choices far in the future may differ from what an individual would prefer if she had an unbiased view of her world. It follows that a better understanding of the influence of various cognitive biases on saving preferences may be important for policy design.

Example 3: Fang and Silverman (2004a,b) and the Design of Welfare Policy.

As with smoking and saving, it is natural to think that time discounting may play an important role in labor supply and welfare program participation decisions. Again, the costs of leaving welfare for work (trading home production, leisure and welfare benefits for a low wage) are immediate, while the benefits, if they exist, are delayed until human capital accumulates and delivers higher wages. Nevertheless, the logic behind investigating the influence of present-biased time preferences on labor supply and welfare program participation decisions would seem to apply to projection bias as well. If individuals tend to project their current tastes onto their future tastes, those currently on welfare may mispredict their future utility from work versus staying home. More specifically, suppose non-working welfare recipients anticipate that leaving children with a baby-sitter, or commuting long distances, or behaving respectfully to a boss will be more distasteful than it, in fact, will be. In that case, even if preferences were time consistent, there may be utility gains from policy interventions that, effectively, create incentives for would-be recipients to enter the paid labor force.

While both projection and present bias may each suggest too much welfare use by some families, precisely how each of these biases affect labor supply and welfare program participation decisions has potential implications for policy design. If, for example, projection bias had an important influence on the decisions of welfare recipients, then programs that promote *gradual* transitions into work and thus permit low cost adaptation of preferences may be more successful in terms of attachments to the labor market than stronger pushes such as workfare or time limits. The opposite may be true if present bias had an important influence on the labor supply of the welfare-eligible. In that case, if future changes in the return to welfare versus work are not abrupt or large then these changes may provide insufficient incentive to overcome the tendency to delay entry into the labor force driven by present bias.

3 Distinguishing Among Biases

Each of the previous examples from the behavioral economics literature suggests that the presence of cognitive biases may have important policy implications, and that optimal policy design may depend on precisely which biases play the most significant roles in decision making. The policy implications of multiple biases motivates the question: How can we determine which biases are most important in which contexts?

As noted above, in some contexts there may be strong *a priori* reasons to think that one bias would play a more important role than another. For example, decisions under uncertainty where costs and rewards are nearly simultaneous, such as the decision to wear a seat-belt or to undertake some medical options, are *a priori* more likely to be influenced by optimism or confirmatory biases than by projection or present biases. In cases like these, researchers could more confidently exclude certain biases from consideration.

Calibration exercises can provide another method for ranking the relative importance of various cognitive biases. Calibration involves developing a model of behavior and choosing parameters of the model, including degrees of bias, so that the behavior it predicts fits summary statistics of relevant data. If there are many combinations of parameters that fit the same data, then the researcher can experiment with different levels of biases and simulate their effects on decision making. With these simulations one can evaluate whether the degree of bias necessary to have an economically substantial impact is plausible. The plausibility of the relevant bias calibrations may, in turn, be evaluated in terms of their implications for other decisions. For example, one may consider whether the degree of projection bias necessary to have an important influence on smoking decisions is compatible with decisions about saving levels.

A third option for determining the most important biases and heuristics for a particular decision domain, is to use field and experimental data to perform direct assessments of the preferences and beliefs that in principle guide decision making.[9] If, for example, both present-bias and optimism bias are logically important contributors to the decision to smoke, survey research could collect data on the beliefs and rates of time discount of both smokers and non-smokers and then assess the ability of these measures to predict behavior. If those with steeper time discount functions are no

[9] Given the difficulties of identifying *any* model of behavior from choices alone, Manski (2004) also argues for the direct measurement of expectations or beliefs to enrich empirical analysis.

more likely to smoke, then present-biased preferences are less likely to be important determinants of behavior. Alternatively, if smokers have no more sanguine beliefs about their expected costs of smoking, then optimism is less like to be a driving force behind the decision to smoke.

Finally, yet another approach to identifying the importance of various biases and heuristics uses economic theory to inform inference based on data about choices. More precisely, economic theory may provide enough structure to allow us to distinguish the effects of various biases from each other, and from a traditional model, using readily available data. In the remainder of this paper, we summarize our investigation into whether, in fact, theory provides a structure sufficient to distinguish the effects of present-bias versus projection bias in the context of the labor supply welfare program participation choices we studied in Fang and Silverman (2004a,b). We begin the summary with an informal description of our model of labor supply and welfare-program participation.

4 The Model

Appendix A presents a formal model of the work-welfare decision faced by a single parent with children. That model incorporates both the present-bias and the projection bias models as special cases. Here we discuss the basic ingredients of the model and outline its relevant predictions.

The model considers the labor supply and welfare program participation choices of a single parent with children. In each period of time,[10] the parent chooses whether to stay at home and receive welfare, work in the labor market, or stay at home and receive no welfare. These choices are mutually exclusive and exhaustive.[11] We normalize the value of staying home without welfare to zero. If the parent instead chooses welfare, she receives a constant real benefit in addition to the value of her leisure and home production (e.g., care-giving and homemaking). The benefit includes both cash payments and in-kind assistance such as health insurance, food stamps and housing subsidies. This is a model without welfare stigma or welfare startup costs; so, in the absence of welfare time limits or work requirement, the parent would never stay home without receiving welfare because doing so would mean forgoing the benefit. If she chooses to enter the labor force, and receives a job offer, her expected wage will depend on how many years she has

[10] We think of a period as six months to a year.

[11] In practice, a single mother can work and be on welfare at the same time, as long as the income is lower than the "break-even" income to retain welfare eligibility. For simplicity, we abstract from this possibility.

worked in the past. This wage captures both money and in kind forms of compensation for work. The relationship between the wage and past work experience reflects the accumulation of human capital, and represents the most important intertemporal link in the model. The opportunity cost of not working now includes the value of future higher wages, which itself depends on how much future market work will be done.

Going to work also involves a utility cost, beyond the lost benefits of forgone welfare payments. This cost of work may have many interpretations: it may include, for example, the opportunity cost of forgone leisure and time with children, the monetary and time cost of child care arrangements, the disutility of keeping to a strict schedule, or the difficulty of familiarizing oneself with public transportation. We assume that this cost of work represents, to some extent, a startup cost. Specifically, the cost of work is lower if the parent worked last period than if she chose either welfare or home last period. The start-up costs may include arranging transportation, developing routines to ensure timely arrival to work, and securing child care services. Such costs will understandably be lower once a routine and support systems have been established. This assumption also reflects the notion that tastes for work will adapt. Agents will simply get used to certain distasteful aspects of labor market work, such as keeping to a strict time schedule, respecting a boss, and dealing with the unreasonable demands of co-workers. We also want to emphasize that the preceding discussion refers the cost of work that will be *actually* experienced by the agent. As described below, we will allow her *perception* of the future cost of work to be biased.

To allow for present-biased preferences, our model follows Laibson (1997) and O'Donoghue and Rabin (1999) who each adopted a relatively simple form of possibly time-inconsistent discounting, so-called (β, δ)-preferences. This is a now standard way of modeling the taste for immediate gratification; details are provided in Appendix A.

Next we describe how the parent forms her expectations for the future that are critical for her current decision making. As described above, there is considerable evidence that decision makers are subject to projection bias when forecasting future tastes. There is little evidence, however, that this bias extends to expectations for the future size of budget sets. We therefore focus in this model on the bias in predicting the disutility of being away from children, or having to keep a strict time schedule, or having to behave respectfully to a boss. That is, we allow projection bias to affect only predictions of the future cost of working. Following Loewenstein, O'Donoghue and Rabin (2003),

we model the projection bias concerning the future cost of working as follows.

When an agent makes predictions about the cost of working in future period, we assume that, to some degree, she projects her current tastes onto her future ones. More precisely, we assume there is no misprediction when the agent is contemplating the cost of working in a future state that, in terms of previous period's decision, is the same as her current situation. But if her current state differs from the one she is contemplating, she predicts her future cost of working will, to some extent, resemble her current cost. Because the costs of work are to some extent start up costs, this latter assumption implies that, when staying at home and receiving welfare, the parent will perceive her cost of working as larger than it will in fact be. The greater her projection bias, the greater her overestimate of the cost of working. Thus our formulation captures in a simple way the idea that when the agent is called upon to predict her taste for work in the future, she overestimates the similarity of her future taste to her current taste.[12]

4.1 Why This Form of Projection Bias?

As explained in the preceding subsection, we have assumed a particular form of projection bias. In what follows, we provide additional justifications for modeling projection bias as a tendency to overestimate the future utility costs of work.

The fact that many welfare recipients lack significant experience in the labor force makes the notion that they mispredict their utility from working more plausible. Just as those who have never lived in a cold climate, and who know few who have, may mispredict their ability to adapt to the winter, so welfare recipients who have never worked, and know few who have, may mispredict their ability to adapt to being away from children, or to commuting, or to showing respect for a boss.

It is not immediately clear, however, that if welfare recipients mispredict their taste for work, they mispredict it with a *negative* bias; they might also overestimate the utility of time spent away from children or the pride in earning a wage. If such positive projection bias were the sole bias influencing welfare decisions then there would be, in some sense, too few welfare recipients

[12] At a more technical level, our model also assumes the parent is "sophisticated" about her projection bias. In this context, sophistication amounts to the agent having the ability to recognize her future tendency to place too much weight on present state preferences, but does not recognize that tendency in her current self. We adopt this assumption of sophistication, in part, to emphasize that the observable differences in present- versus projection-biased decision makers do not depend on having different levels of sophistication about these biases.

and restrictions aimed at moving women from welfare to work would certainly make the former recipients worse off. We are not aware of direct empirical evidence regarding the accuracy of the predictions of welfare recipients about their utility of work.[13] However, an Urban Institute study of mothers in 1997 found that, conditional on demographic and economic characteristics, welfare recipients were less likely than other mothers to agree with the statement that a working mother can establish just as warm a relationship with her children as a mother who does not work (Wertheimer, et al. 2001). These differences in attitudes about work and parenting represent the opinions of different women in different work-welfare situations, and not the opinions of the same women at different times and in different situations. Thus the findings are consistent with a simple model of selection into work based on preferences for work. These finding are also consistent, however, with welfare recipients being subject to negative projection bias about their tastes for work.

More generally, suppose we take a traditional approach and think of time spent away from work as a desirable good. In that case assuming that parents underestimate their ability to adjust to less of that good is consistent with evidence from a large literature on hedonic adaptation. That literature shows, quite robustly and across a wide range of domains, that individuals overestimate the disutility that unpleasant circumstances will give them over the long term.[14] We view this evidence of individuals' systematic underestimation of their ability to adapt to negative circumstances (along with the conventional treatment of leisure as a good) as favoring the modelling choice of a negative projection bias regarding predictions of the future taste for work.

4.2 Projection Bias Model vs. Present Bias Model

Our most general model allows any combination of projection bias and hyperbolic discounting models. However, our formal analysis limits attention to two classes of models that are nested as special cases of the general model outlined above:

- **Projection Bias Model:** If the general model is restricted to have time-consistent preferences, but continues to allow projection bias, we will call it the projection bias model;

[13] We emphasize the mispredictions considered here are limited to biased beliefs about future tastes. We do not, for example, consider mispredictions about the size of future budget sets. There is relatively little evidence that individuals systematically mispredict how changes in circumstance will affect their budgets for consumption or time.

[14] See, e.g., Sackett and Torrance (1978), Gilbert *et al* (1998), and Riis *et al* (2004).

- **Present Bias Model:** If the above model is restricted to allow no projection bias, but continues to permit present bias, we will call it the present bias model.

In what follows we summarize our analysis of whether the projection bias and present bias models can be distinguished using data on choices, outcomes and attitudes toward welfare policies.

5 Distinguishing Biases Using Data on Choices, Outcomes and Attitudes

5.1 Choice and Outcome Data

To establish that one can empirically distinguish the projection bias models and present bias models using only data on choices and possibly outcomes (e.g., wages), one has to show formally that the observed data is consistent with one model but not the other. In Appendix B, we conduct such an exercise and show that, indeed, standard data can be used to distinguish a present-biased individual from a decision maker with projection bias, and each from a time-consistent, rational decision maker. More precisely, in the context of the models presented in Section A, there is no set of parameters such that the behavior of a present-biased agent could be replicated by that of an agent with projection bias. Similarly, the behavior of a time-consistent agent cannot be replicated by an agent with either bias.

Why can't the optimal behavior and outcomes of a population of present-biased parents be replicated by the optimal behavior of either time-consistent or projection-biased parents, and vice versa? The analysis in Appendix B shows that such replication is impossible because the optimizing choices and outcomes of present- and projection-biased agents must, at some point, be inconsistent. In particular, data on the welfare benefit level and the lowest wages ever accepted at the very end of the planning horizon allow us to pin down the common costs of working across all three models. Similarly, the probabilities of various choices, conditional on work experience, wages accepted and welfare benefits pin down the values that are derived from different various choices, regardless of the underlying model that is driving that behavior. With the same costs, same values, and same outcomes (in particular the same wages), the equations describing optimizing behavior in various states of the world cannot be simultaneously satisfied for both present-biased and projection-biased

populations. In this setting, the data can support one interpretation or the other, but not both. A similar argument distinguishes an unbiased population from either a present-biased or projection-biased population.

These results are useful to the extent that they inform us about the driving forces for identification. However, such identification results rely crucially on the structures imposed by the models; as such they should be interpreted with care. Specifically, we show that if the true data generating process is a present bias model in the form of (quasi-) hyperbolic discounting, then such data cannot be rationalized by a model of a projection bias as modeled in the simple form found in Loewenstein, O'Donoghue and Rabin (2003). That is, we are able to distinguish the quasi-hyperbolic discounting from a particular model of simple projection bias. To the extent that quasi-hyperbolic discounting and simple projection bias models are specifications designed for analytic convenience and do not capture present and projection biases more generally, the applications of our result are limited. It may, for example, be impossible to distinguish between projection bias and a more general form of present-biased time preferences. Caution in interpreting these results is also justified because we do not know yet, if the true data generating process is a model with a combination of present and projection biases, whether we will be able to disentangle the magnitude of these biases from the standard data.

5.2 Attitude Data

In part because the technical argument for distinguishing between biases relies on specific formulations of those biases, we also considered a reduced-form method of distinguishing between these models based on the idea that agents with projection bias will often have different attitudes toward changes in welfare policy such as time limits or work requirements from those of present-biased agents. In particular, the change in attitude toward a policy that an agent exhibits once the policy is actually implemented may be different for projection-biased versus present-biased agents. These differences in changes in attitude can, in principle, be measured and thus distinguish behavior driven by present bias versus projection bias.

Appendix C presents the details of the formal analysis. That analysis shows that there exists an empirically plausible set of circumstances under which the attitudes of the welfare-eligible to time limit policies could be used to distinguish among unbiased, present-biased and projection-biased

parents. A critical ingredient of these circumstances is that net wages must at some point decline with experience and then increase. Under these circumstances, prolonged welfare program participation for women with present-biased preferences is driven, in the terms of Fang and Silverman (2004a), by a "lack of commitment" to work through the difficult times. By "lack of commitment," we mean that an individual would strictly prefer to work now *if* she could commit herself to continue working in the future; however, due to the self-control problems that derive from present-biased preferences, she knows that in the absence of an external commitment device she would not, in fact, work in the future, even if she worked today. The basic insight is that would-be welfare recipients with projection bias do not want time limits or workfare ex ante, but may prefer it ex post, when they realize the costs of working are not as high as they had earlier thought. Welfare recipients with present bias may want time limits or workfare ex ante, but will at some point prefer to relax it ex post, as they fall victim to the desire for immediate gratification. We summarize these results as two hypotheses:

Hypothesis 1: If the early selves of present-biased agents choose welfare as a result of "lack of commitment," then such agents will exhibit favorable attitudes toward time limits prior to the implementation of time limits; however, once time limits are actually implemented, they will prefer that time limits be relaxed.

Hypothesis 2: If agents have projection bias about the cost of working, then they will exhibit negative attitudes toward time limits prior to the implementation of time limits, but once time limits are actually implemented, they will prefer that time limits remain in place.

These two hypotheses form the basis of an empirical strategy. They suggest that, if we have accurate measures of the *attitudes* toward time limits before and after the limits are implemented, we can distinguish the present bias from the projection bias. An important question is, to what extent can we rely on attitudinal data from surveys? In contrast to researchers in other social sciences, economists have traditionally shied away from subjective attitudinal data, and instead relied on behavioral data to make statistical inferences. Common complaints about subjective attitudinal data include:[15]

[15] Bertrand and Mullainathan (2001) has an interesting discussion about the reliability of subjective survey data.

1. survey respondents do not have incentive to think hard about the question;

2. their responses are very sensitive to the wording of the question;

3. people simply do not have opinions about many things and if you force them to form an opinion, their answers will be uninformative.

The first complaint is applicable to any retrospective survey data, including commonly used behavioral data such as Current Population Survey. The second and third complaints are more relevant. Indeed researchers have found that slight changes in the framing of the questionnaire can have big effects on how respondents answer these questions. These valid concerns, however, should not be the reason for not using attitudinal data; rather it calls for careful questionnaire design, and the adoption of attitude elicitation techniques from cognitive psychology (see Manski 2004).[16]

6 Conclusion

Recent research that incorporates cognitive biases into economic models of decision making typically studies the implications of these deviations from the standard model one bias at a time. This strategy of limiting attention to one potential bias may, in some cases, be justified by a priori indications of the primacy of that bias in determining behavior. If, however, several biases might plausibly explain behavior, the obvious question for empirical research is whether data allow these various biases to be distinguished from each other, and from a traditional model. Distinguishing between biases is particularly important when the policy implications of the underlying explanations for behavior differ substantially. In this paper, took up this question for the case of labor supply and welfare program participation.

We analyzed a simple model of work and welfare program participation with two possible psychological biases, time-inconsistent discounting in the form present-biased time preferences and non-rational beliefs in the form of projection bias; and explored the possibilities to distinguish these two biases. We showed, formally, that we can distinguish between models of present and

[16] More and more economists are now using subjective survey data to answer important policy questions. For example, Shapiro and Slemrod (2003) used consumer surveys to study the extent to which households spend the tax rebate in 2001; Gruber and Mullainathan (2002) use self-reported happiness to measure the welfare consequence of cigarette tax hike.

projection bias, and each from a standard model, using only standard data. In addition, if present-biased agents choose to participate in welfare instead work as a result of a "lack-of-commitment," then they would exhibit a favorable attitude toward time limits or work requirements before the implementation of such eligibility restrictions, but would prefer them to be relaxed once they are implemented. In contrast, single mothers with projection bias would ex ante be against eligibility restrictions, but will be in favor of them once they are actually imposed. We argued that to the extent that attitudes toward policy changes can be accurately elicited and measured, the above ideas can be exploited to distinguish behavior driven by present bias versus behavior driven by projection bias. The ideas in this paper are exploratory; they represent a small step in an important direction of examining how the increasing number of cognitive biases that economists are introducing into economic models can be potentially distinguished.

A Formal Model

In this appendix we present a formal model of labor supply and welfare program participation decisions that incorporates both present and projection biases.[17] Consider a discrete time, finite horizon model with periods $t \in \{1, ..., T\}$. In each period, an agent chooses either to receive welfare (option 0), or work in the labor market (option 1), or stay at home without work or welfare (option 2). An agent's choice set is denoted by $\mathcal{D} \equiv \{0, 1, 2\}$. Her period-$t$ choice is denoted by $d_t \in \mathcal{D}$.

If the agent chooses **welfare**, she receives a benefit $b > 0$. If the agent receives a job offer and chooses **work**, her *expected* wage depends on the cumulative number of periods she has ever worked, denoted by τ. The average wage as a function of experience is denoted by $\omega(\tau)$. The wage offer in period t, w_t, for an agent with experience τ is

$$w_t = \omega(\tau) + \eta_t$$

where η_t is a mean zero residual drawn from a continuous CDF $G_\tau(\cdot)$ that may depend on τ.

When an agent chooses work, she incurs a direct utility cost, c_t. This *actual* period-t cost of working depends on her period-$(t-1)$ choice, d_{t-1}, and satisfies:

$$c_t = c(\tau, d_{t-1}) = \begin{cases} c_l(\tau) & \text{if } d_{t-1} = 1 \\ c_h(\tau) & \text{if } d_{t-1} \in \{0, 2\} \end{cases} \tag{A1}$$

where $0 < c_l(\tau) < c_h(\tau)$.Thus we assume the cost of continued work is lower than the cost upon transitioning into work from either home or welfare. We emphasize that c_t is the experienced cost of work. We allow the agent's *perception* of her future cost of work to be biased.

In period t an agent's job offer probability, ρ_t, depends on both her work experience and her period-$(t-1)$ choice, d_{t-1}, as follows:

$$\rho_t = \rho(\tau, d_{t-1}) = \begin{cases} \rho_h(\tau) & \text{if } d_{t-1} = 1 \\ \rho_l(\tau) & \text{if } d_{t-1} \in \{0, 2\} \end{cases}$$

where $\rho_h(\tau) \in (0, 1), \rho_l(\tau) \in (0, 1), \rho_h(\tau) \geq \rho_l(\tau)$ and $\rho_h(\cdot)$ and $\rho_l(\cdot)$ are increasing in τ. These assumptions capture the idea that offers are easier to get when working, and with more experience.

Home. We normalize the payoff from staying home without welfare to zero and assume all agents begin the decision process at home without welfare. In the absence of time limits or work requirements, the choice of home without welfare is dominated by welfare.

[17] This model is an extension of Fang and Silverman (2004a).

State Variable and Strategies. When an agent makes her choice in period t, her relevant state variable is $h_t \equiv (t, \tau, d_{t-1}, \chi_t, \eta_t)$, indicating that the period is t; she has worked τ periods; and her decision last period was $d_{t-1} \in \mathcal{D}$. The variable $\chi_t \in \{0, 1\}$ indicates whether the agent receives the offer $\omega(\tau) + \eta_t$ in period t. We will often refer to the second element of h_t as τ_t.

We restrict attention to Markovian strategies; a *feasible strategy* in period t, σ_t, is a mapping from the set of all possible period t states into the choice set $\mathcal{D}$. Given h_t and σ_t, the state variable in period $t+1$ is denoted $\overrightarrow{h}_{t+1}(h_t, \sigma_t)$. A *strategy profile* is a vector of mappings $\boldsymbol{\sigma} \equiv (\sigma_t)_{t=1}^{T}$, specifying for each period the agent's action in all possible states. For any period $s > t$, we denote the period-s state that will be reached from h_t if the strategy profile $\boldsymbol{\sigma}$ is followed by $\overrightarrow{h}_s(h_t, \boldsymbol{\sigma})$.

Actual Instantaneous Utility Function. We assume that the agent cannot borrow or save. Thus her period-t instantaneous utility is given by

$$u_t = u(d_t, h_t) = \begin{cases} b & \text{if} \quad d_t = 0 \\ w_t - c(\tau, d_{t-1}) & \text{if} \quad d_t = 1 \\ 0 & \text{if} \quad d_t = 2, \end{cases} .$$

A.1 Modeling Projection Bias

We assume projection bias affects only predictions of the future cost of working. Borrowing from Loewenstein *et al.* (2003), we model this projection bias as follows. In period t, the agent's predicted cost of working in period $s \geq t+1$, with experience τ_s, denoted $\tilde{c}_s$, satisfies:

$$\begin{aligned} \tilde{c}_s &= \tilde{c}(\tau_s, d_{s-1} | d_{t-1}) \\ &= (1-\alpha) c(\tau_s, d_{s-1}) + \alpha c(\tau_s, d_{t-1}), \end{aligned} \tag{A2}$$

where $\alpha \in [0, 1]$. Note that, if $d_{s-1} = d_{t-1}$, then $\tilde{c}_s = c_s$. That is, there is no misprediction when the agent is contemplating the cost working in a future state that, in terms of previous period's decision, is the same as her current situation. The parameter $\alpha \in [0, 1]$ reflects the *degree of projection bias*. When $\alpha = 0$, the agent accurately predicts her cost of working; when $\alpha = 1$, the agent perceives that future cost of working will be what her current cost of working would be with experience τ_s. When $\alpha \in (0, 1)$, the agent has an intermediate level of projection bias.

Projected Future Utilities. Reflecting this potentially biased perception of future utility,

the instantaneous period-s utility, projected by the agent in period $t \leq s$ is

$$\tilde{u}_s\left(d_s, h_s | d_{t-1}\right) = \begin{cases} b & \text{if} \quad d_s = 0 \\ w_s - \tilde{c}\left(\tau_s, d_{s-1} | d_{t-1}\right) & \text{if} \quad d_s = 1 \\ 0 & \text{if} \quad d_s = 2, \end{cases} .$$

A.2 Intertemporal Preferences and Present Biased Discounting

We represent an agent's possibly time-inconsistent preferences with so-called (β, δ)-preferences. For all t, if an agent predicts a stream of utilities from period t onward $\left(\tilde{u}_t, ..., \tilde{u}_T\right)$, then the agent will evaluate that stream according to

$$V^t\left(\tilde{u}_t, ..., \tilde{u}_T | d_{t-1}\right) \equiv \delta^t \tilde{u}_t + \beta \sum_{s=t+1}^{T} \delta^s \tilde{u}_s, \text{where } \beta \in (0,1], \delta \in (0,1]. \tag{A3}$$

The parameter δ reflects long-run, time-consistent discounting and is called the *discount factor*. The parameter β captures short-term impatience, and is called the *present bias factor*. When $\beta = 1$, (β, δ)-preferences are time-consistent. When $\beta \in (0,1)$,preferences are present-biased and time-inconsistent.

Following a standard approach, we analyze the behavior of an agent by thinking of the decision-maker in each period as a separate *self*. Each period-t self makes her choice to maximize her current preferences $V^t\left(\tilde{u}_t, ..., \tilde{u}_T | d_{t-1}\right)$, knowing her future selves control their own behavior.

A.3 Sophistication and Equilibrium

The literature on time-inconsistent preferences distinguishes between *sophistication* and *naivete* (Strotz 1955, Pollak 1968, O'Donoghue and Rabin 1999). An agent is *sophisticated* if each period-t self knows her future selves' preferences and anticipates their behavior when making her own choice. She is *naive* if each period-t self believes that her future selves' preferences are identical to her own. We assume agents are sophisticated with respect to both their present and projection biases. See the main text for a discussion.

Our equilibrium concept is an analog of subgame perfection for the intrapersonal game and generalizes "perception-perfection" from O'Donoghue and Rabin (1999). Given a strategy profile $\boldsymbol{\sigma}$, denote by $\tilde{W}_s\left(\boldsymbol{\sigma}, h_s | d_{t-1}\right)$ the *expected continuation long-run* utility from period s $(s \geq t+1)$

onward *perceived* by the period-t self. This value is defined recursively from period T as follows:

$$\tilde{W}_T\left(\boldsymbol{\sigma}, h_T|d_{t-1}\right) = \tilde{u}_T\left(\sigma_T\left(h_T\right), h_T|d_{t-1}\right), \tag{A4}$$

and for $t = T-1, ..., t+1$,

$$\tilde{W}_s\left(\boldsymbol{\sigma}, h_s|d_{t-1}\right) = \tilde{u}_s\left(\sigma_s\left(h_s\right), h_s|d_{t-1}\right) + \delta\mathbb{E}\tilde{W}_{s+1}\left(\boldsymbol{\sigma}, \overrightarrow{h}_{s+1}\left(\sigma_s\left(h_s\right), \boldsymbol{\sigma}\right)|d_{t-1}\right).$$

In words, this value represents the period-t self's intertemporal preferences from some prior perspective where her own present bias is irrelevant, but accounting for her projection bias.

Definition 1 *A perception-perfect equilibrium for an agent with potentially both present and projection biases is a strategy profile $\boldsymbol{\sigma}^* \equiv \{\sigma_t^*\}_{t=1}^T$ that satisfies, for all h_t, and for all t,*

$$\sigma_t^*\left(h_t\right) = \arg\max_{d_t \in \mathcal{D}} \left\{u_t\left(d_t, h_t\right) + \beta\delta\mathbb{E}\tilde{W}_{t+1}\left(\boldsymbol{\sigma}^*, \overrightarrow{h}_{t+1}\left(h_t, d_t\right)|d_{t-1}\right)\right\}.$$

If an agent has no projection bias, our equilibrium concept reduced to that used in O'Donoghue and Rabin (1999). The analysis that follows restricts attention to two classes of models that are nested as special cases of our general model. If the general model is restricted to $\beta = 1$, we call it the **projection bias model**. When we assume $\alpha = 0$, we call it the **present bias model**.

B Distinguishing Projection and Present Bias in Standard Data

In this appendix we show that the present bias and projection bias models can distinguished in standard data, and that each can be distinguished from a traditional model with no biases.

First we make precise what it means to say two models can be distinguished, or that a model is (partially) identified. Model A can be *distinguished from* model B if there exists no parameterization of model B that is consistent with a large sample of data if those data were actually generated by model A, and vice versa. We say that a model is *(partially) identified* if we can uniquely recover from the data all (some) of the primitives of the true data-generating model.[18].

Standard Data. In what follows, we assume access to a data set consisting of an infinite number of individuals with observations on:

[18] Questions of distinguishing or identifying a model are different from questions of estimation. A study of distinction or identification assumes an infinitely large sample and abstracts from the difficulties posed by finite sample estimation.

[DA]. The experience level and choices for all individuals i at each period t: $\left\{\tau_t^i, d_t^i\right\}_{t=1,...,T}^{i\in\mathcal{I}}$;

[DB]. The welfare benefit level b;

[DC]. The accepted wages of those who work $\left\{w_t^i\right\}_{t=1,...,T}^{i\in\mathcal{I}}$, if $\chi_t^i = 1$ and $d_t^i = 1$;

To determine whether we can distinguish the projection and present bias models, we begin by assuming the data were generated by a present bias model with unknown primitives (denoted with a superscript *). We then ask whether there exists a projection bias model $\left(\beta' = 1, \delta', \alpha' \in (0,1)\right)$ that could have generated the data that was actually generated by the present bias model $(\beta^*, \delta^*, \alpha^* = 0)$.

Our first result shows that we can recover the working cost functions from the standard data:

Lemma B1 *We can recover $c_l(\tau)$ and $c_h(\tau)$ from **[DB]** and **[DC]**.*

Proof. In period T, given any state h_T, the decision rule is simple: the agent works if and only if

$$\omega(\tau) + \eta_T \geq c(\tau, d_{T-1}) + b.$$

With infinite data, **[DC]** gives the lowest acceptable wage for an agent with experience τ_T, whose prior choice was d_{T-1}. Let $\underline{w}_T\widehat{(\tau, d_{T-1})}$ denote the lowest wage among those who chose d_{T-1} in the previous period. Consistent estimates of $c_l(\tau)$ and $c_h(\tau)$ are then given by

$$\begin{aligned} \widehat{c_l(\tau)} &= \underline{w}_T\widehat{(\tau, 1)} - b, \\ \widehat{c_h(\tau)} &= \underline{w}_T\widehat{(\tau, 0)} - b. \blacksquare \end{aligned} \tag{B5}$$

Proposition 1 *If $T \geq 3$, then projection bias and present biased models are nonparametrically distinguishable in standard data.*

Proof. From **[DA]**, we can consistently estimate the conditional choice probabilities at date t, denoted by $\Delta_t(d_t|\tau, d_{t-1})$, from their sample analogues. For example,

$$\widehat{\Delta_t(1|\tau, d)} = \frac{\#\left\{i : d_t^i = 1, \tau_t^i = \tau, d_{t-1}^i = d\right\}}{\#\left\{i : \tau_t^i = \tau, d_{t-1}^i = d\right\}}. \tag{B6}$$

Similarly, we can estimate the average wages for working agents in period t conditional on experience τ and d_{t-1}. They are denoted by $\bar{I}_t(\tau, d_{T-1})$, and estimated consistently by:

$$\widehat{\bar{I}_t(\tau, d)} = \frac{\sum\limits_{\left\{i : d_t^i = 1, \tau_t^i = \tau, d_{t-1}^i = d\right\}} w_t^i}{\#\left\{i : d_t^i = 1, \tau_t^i = \tau, d_{t-1}^i = d\right\}}.$$

With these estimated conditional choice probabilities and average earnings, we can use the approach of Hotz and Miller (1993) to estimate expected continuation value functions. Consider the terminal period T. Denote the expected continuation value of an agent with experience τ who made period-$(T-1)$ choice d_{T-1} by $Q_T(\tau, d_{T-1})$. Consistent estimates of those expected continuation values are given by:

$$\widehat{Q_T(\tau,d)} = \widehat{\Delta_T(1|\tau,d)}\left[\widehat{\bar{I}_T(\tau,d)} - \widehat{c_h(\tau)}\right] + \left[1 - \widehat{\Delta_T(1|\tau,d)}\right] b \ .$$

Now suppose that a projection bias model could also generate the data generated from this present bias model. In the present bias model (β^*, δ^*), the choice of an agent with experience τ in period-$(T-1)$ is determined as follows. If $d_{T-2} = 0$, then she works only if

$$\omega(\tau) + \eta_{T-1} - c_h(\tau) + \beta^*\delta^* Q_T(\tau+1, 1) \geq b + \beta^*\delta^* Q_T(\tau, 0) \, .$$

Thus the lowest accepted wage in period-$(T-1)$ for agents with experience τ and $d_{T-2} = 0$ is

$$\underline{w}^*_{T-1}(\tau, 0) = b + c_h(\tau) + \beta^*\delta^*\left[Q_T(\tau, 0) - Q_T(\tau+1, 1)\right]. \tag{B7}$$

Similarly, the lowest accepted wage in period-$(T-1)$ for agents with experience τ and $d_{T-2} = 1$ must satisfy:

$$\underline{w}^*_{T-1}(\tau, 1) = b + c_l(\tau) + \beta^*\delta^*\left[Q_T(\tau, 0) - Q_T(\tau+1, 1)\right]. \tag{B8}$$

Under a *projection bias* model $\left(\delta', \alpha'\right)$, an agent with experience τ and $d_{T-2} = 0$ will work in period-$(T-1)$ only if

$$\omega(\tau) + \eta_{T-1} - c_h(\tau) + \delta'\tilde{Q}_T(\tau+1, 1|d_{T-2}=0) \geq b + \delta'\tilde{Q}_T(\tau, 0|d_{T-2}=0)$$

where $\tilde{Q}_T$ is the net continuation value at period T perceived by the period-$(T-1)$ agent. Note that $\tilde{Q}_T(\tau, 0|d_{T-2}=0) = Q_T(\tau, 0)$,and,

$$\begin{aligned}
\tilde{Q}_T(\tau+1, 1|d_{T-2}=0) &= \Delta_T(1|\tau+1, 1)\left\{\bar{I}_T(\tau+1, 1) - \left[\left(1-\alpha'\right)c_l(\tau+1) + \alpha' c_h(\tau+1)\right]\right\} \\
&\quad + \left[1 - \Delta_T(1|\tau+1, 1)\right] b \\
&= Q_T(\tau+1, 1) - \alpha'\Delta_T(1|\tau+1, 1)\left[c_h(\tau+1) - c_l(\tau+1)\right].
\end{aligned}$$

Thus the lowest accepted wage in period-$(T-1)$ for agent with experience τ and $d_{T-2} = 0$ is

$$\begin{aligned}
\underline{w}'_{T-1}(\tau, 0) &= b + c_h(\tau) + \delta'\left[Q_T(\tau, 0) - Q_T(\tau+1, 1)\right] \\
&\quad + \alpha'\delta'\Delta_T(1|\tau+1, 1)\left[c_h(\tau+1) - c_l(\tau+1)\right]
\end{aligned} \tag{B9}$$

An analogous argument for an agents with experience τ and $d_{T-2} = 1$ shows that their lowest accepted wage in period-$(T-1)$ must satisfy

$$\begin{aligned} \underline{w}'_{T-1}(\tau, 1) &= b + c_l(\tau) + \delta' \left[Q_T(\tau, 0) - Q_T(\tau + 1, 1) \right] \\ &\quad + \alpha' \delta' \Delta_T(1|\tau, 0) \left[c_h(\tau) - c_l(\tau) \right]. \end{aligned} \tag{B10}$$

If the two models are indistinguishable, it must be that

$$\begin{aligned} \underline{w}'_{T-1}(\tau, 0) &= \underline{w}^*_{T-1}(\tau, 0) \\ \underline{w}'_{T-1}(\tau, 1) &= \underline{w}^*_{T-1}(\tau, 1). \end{aligned}$$

The costs of work, $c_h(\tau)$ and $c_l(\tau)$, are identified from Lemma 1; they cannot be chosen differently for different models. Thus the two preceding wage equalities are satisfied only if

$$\begin{aligned} \beta^* \delta^* \left[Q_T(\tau, 0) - Q_T(\tau + 1, 1) \right] &= \delta' \left[Q_T(\tau, 0) - Q_T(\tau + 1, 1) \right] \\ &\quad + \alpha' \delta' \Delta_T(1|\tau + 1, 1) \left[c_h(\tau + 1) - c_l(\tau + 1) \right] \\ \beta^* \delta^* \left[Q_T(\tau, 0) - Q_T(\tau + 1, 1) \right] &= \delta' \left[Q_T(\tau, 0) - Q_T(\tau + 1, 1) \right] \\ &\quad + \alpha' \delta' \Delta_T(1|\tau, 0) \left[c_h(\tau) - c_l(\tau) \right]. \end{aligned}$$

These two equalities can be simultaneously satisfied only if $\delta' = \beta^* \delta^*$ and $\alpha' = 0$. Proposition 2, below, goes on to show that if $T \geq 3$, then the present bias model (β^*, δ^*) can also be distinguished from an exponential discounting model. Thus the projection and present bias models can be distinguished from each other and each can be distinguished from a standard model. ■

Proposition 2 *If $T \geq 3$, then a present bias model with $\beta^* \in (0, 1), \delta^* \in (0, 1)$ can be distinguished from an exponential discounting model with $\beta' = 1$ and $\delta' \in (0, 1)$ using standard data.*

Proof. Following the proof of the previous proposition, an unbiased model rationalizes the minimum accepted wage at period-$(T-1)$ only if $\delta' = \beta^* \delta$.

We can estimate the period-$(T-1)$ expected continuation value of an agent with experience τ and period-$(T-2)$ choice d_{T-2} as $Q_{T-1}(\tau, d_{T-2})$. Consistent estimates of these expected continuation values for a (β^*, δ^*) agent are given by:

$$\begin{aligned} \widehat{Q_{T-1}(\tau, d)} &= \widehat{\Delta_{T-1}(1|\tau, d)} \left[\widehat{\bar{I}_{T-1}(\tau, d)} - \widehat{c_h(\tau)} + \delta^* \widehat{Q_T(\tau + 1, 1)} \right] \\ &\quad + \left[1 - \widehat{\Delta_{T-1}(1|\tau, d)} \right] \left[b + \delta^* \widehat{Q_T(\tau, 0)} \right] \end{aligned}$$

Now consider an agent with (β^*, δ^*) in period-$(T-2)$. Analogous to the argument in the proof of Proposition 1, the lowest accepted wage in period-$(T-2)$ for agents with experience τ and $d_{T-3} = 0$ must be

$$\underline{w}^*_{T-2}(\tau, 0) = b + c_h(\tau) + \beta^*\delta^* \left[Q_{T-1}(\tau, 0) - Q_{T-1}(\tau + 1, 1) \right].$$

An analogous argument for a exponential discounting agents with $\delta' = \beta^*\delta^*$ shows that their lowest accepted wage in period-$(T-2)$ for agents with experience τ and $d_{T-3} = 0$ is

$$\underline{w}'_{T-2}(\tau, 0) = b + c_h(\tau) + \delta' \left[Q'_{T-1}(\tau, 0) - Q'_{T-1}(\tau + 1, 1) \right].$$

Note $\underline{w}^*_{T-2}(\tau, 0)$ and $\underline{w}'_{T-2}(\tau, 0)$ weight $\widehat{Q_T(\tau + 1, 1)}$ and $\widehat{Q_T(\tau, 0)}$ differently unless $\beta^* = 1$. ■

So far, we have shown that, in standard data, we can distinguish projection bias and present bias models; and distinguish a present bias model ($\beta^* < 1$) from a standard model. These results do not imply, however, that we can identify all the primitives of these models from standard data. Under some distributional restrictions, however, we can guarantee at least partial identification.

Proposition 3 *We can recover $\omega(\tau)$ for all $\tau > 0$ up to a constant from standard data if the following two conditions are satisfied:*

1. $G_\tau(\cdot) = G(\cdot)$;
2. $g'(\eta) \equiv \partial^2 G(\eta)/\partial\eta^2 \neq 0$ *almost everywhere.*

Proof. For this proof, we will only use data from period T. Write the period-T cumulative accepted wage distribution of agents with experience τ and $d_{T-1} = 0$ as $F^\tau_T(w)$. From such agents' period-T optimal decision rule (see the expression in the proof of Lemma B1), we have that, if $w \geq c_h(\tau) + b$,

$$\begin{aligned} F^\tau_T(w) &= \Pr\left[\omega(\tau) + \eta \leq w | \omega(\tau) + \eta - c_h(\tau) \geq b, \chi_T = 1\right] \\ &= \Pr\left[\eta \leq w - \omega(\tau) | \eta \geq c_h(\tau) + b - \omega(\tau), \chi_T = 1\right] \\ &= \frac{G(w - \omega(\tau)) - G(c_h(\tau) + b - \omega(\tau))}{\left[1 - G(c_h(\tau) + b - \omega(\tau))\right] \rho_l(\tau)} \\ &= \frac{G(w - \omega(\tau)) - G(c_h(\tau) + b - \omega(\tau))}{\Delta_T(1|\tau, 0)}; \end{aligned}$$

and $F^\tau_T(w) = 0$ otherwise. The last equality holds because $\left[1 - G(c_h + b - \omega(\tau))\right] \rho_l(\tau)$ is simply the period-T probability of working for agents with experience τ and $d_{T-1} = 0$, which is observable

in the data [see formula (B6) for its empirical analogue]. Thus the period-T accepted wage density of agents with experience τ and $d_{T-1}=0$ is

$$f_T^{\tau}(w)=\begin{cases}\frac{g(w-\omega(\tau))}{\Delta_T(1|\tau,0)} & \text{if } w\geq c_h(\tau)+b\\ 0 & \text{otherwise.}\end{cases} \tag{B11}$$

From **[DC]**, we could observe the empirical analogue of $f_T^{\tau}(w)$ for all $\tau\geq 0$. Denote these empirical counterparts by $\widehat{f_T^{\tau}(w)}$. Together with $\widehat{\Delta_T(1|\tau,0)}$ from formula (B6), we write

$$\widehat{h_T^{\tau}(w)}=\widehat{f_T^{\tau}(w)}\widehat{\Delta_T(1|\tau,0)} \text{ if } w\geq c_h(\tau)+b.$$

Now using (B11), we can construct estimates of the upper tails of $g(\eta)$. For example, from agents with experience $\tau=0$ and $d_{T-1}=1$, we have

$$\widehat{g(\eta)}=h_T^0\left(\widehat{\eta+\omega}(0)\right) \text{ if } \eta\geq c_h(0)+b-\omega(0). \tag{B12}$$

From agents with experience $\tau=1$ and $d_{T-1}=1$, we have

$$\widehat{g(\eta)}=h_T^1\left(\widehat{\eta+\omega}(1)\right) \text{ if } \eta\geq c_h(1)+b-\omega(1). \tag{B13}$$

Assumption 2 ensures that the upper tail of $g(\eta)$ can be identified. Thus we know (B12) and (B13) have to coincide for $\eta\geq\max\{c_h(0)+b-\omega(0), c_h(1)+b-\omega(1)\}$. That is, if we shift $\widehat{h_T^1(w)}$ or $\widehat{h_T^0(w)}$, their right tails have to exactly overlap because each estimates the upper tail of $g(\eta)$. The shift required for the right tails of $\widehat{h_T^1(w)}$ or $\widehat{h_T^0(w)}$ to overlap is exactly the difference between $\omega(0)$ and $\omega(1)$. Thus $\omega(1)-\omega(0)$ is identified. Similarly, $\omega(\tau)-\omega(0)$ is identified for all $\tau>1$.■

Proposition 3 shows that $\omega(\tau)$ can be identified up to a constant under some (non-parametric and testable) distributional restrictions on $G_\tau(\cdot)$; still $G(\cdot)$ itself is not identified. In fact, since we only observe data with information about the upper tails of G_τ, standard data would never allow us to identify the whole distributions of G_τ in the absence of parametric assumptions. If G_τ is not identified, then $\rho_h(\tau)$ and $\rho_l(\tau)$ are not be identified. Finally note that for the purpose of distinguishing projection and present bias models, it is not necessary to know $\rho_h(\tau)$ and $\rho_l(\tau)$.

Cautionary Note: The identification results we have just derived apply to a model in which present biased preferences take the form of quasi-hyperbolic discounting and the projection bias takes the simple form from Loewenstein, *et al.* (2003). To the extent that these specifications are adopted for analytic convenience and do not capture present and projection biases more generally,

the applications of our result are limited. Moreover, we do not know yet, if the true data generating process is a model with a combination of present and projection biases, whether we will be able to disentangle the magnitude of these biases from standard data.

C Changing Attitudes

In this appendix we demonstrate bias-dependent changes in attitude about welfare policies in simple three period examples that are special cases of the general model presented in Appendix **??**.

Suppose the wage-experience profile is deterministic ($\eta_t \equiv 0$), and is given by the following:

τ	0	1	2
$w(\tau)$	0	-9	23.5

The important feature of this wage profile is that net wages get worse with experience before they get better. See Fang and Silverman (2004a) for a more detailed discussion of such a profile. In addition, suppose, for simplicity, that:

- Welfare benefit level is constant at $b = 1$;
- Job offers are always available ($\rho_h = \rho_l = 1$);
- The costs of working are $c_h = 4, c_l = 1$;
- For a present biased agent: $\alpha = 0, \beta = 1/2, \delta = 1$;
- For a projection biased agent: $\alpha = 1, \beta = 1, \delta = 1$.

Suppose that in period 0 agents find themselves out of work. We will analyze their attitudes toward a proposed one-period welfare eligibility time limit and show that these attitudes are different for a present-biased versus a projection-biased agent.

C.1 Attitudes of a Present-biased Agent Toward a Time Limit

Consider first the optimal decisions of a present-biased agent in the absence of a time limit. In period 3, if $\tau_3 = 2$, she must have worked in period 2. In this case, the agent chooses work since

$$w(2) - c_l = 23.5 - 1 > b = 1.$$

If, instead, period two was her first time in the labor market then the agent does not work since

$$w(1) - c_l = -9 - 1 < b = 1$$

Similarly if $\tau_3 = 1$ and $d_2 = 0$, the agent does not work, and if the agent has not yet worked she will not work in the third period. Thus in period 3, an agent works if and only if $\tau_3 = 2$.

Working backwards, consider her optimal plan in period two. Even if she worked in period 1, $\tau_2 = 1$, she does not work in period 2 because

$$w(1) - c_l + \beta [w(2) - c_l] = -9 - 1 + \frac{1}{2}[23.5 - 1] = 1.25 < b + \frac{1}{2}b = 1.5.$$

Similarly, if $\tau_2 = 0$ the agent does not work because $w(0) - c_h < b$ and she would not work in the third period. In short, this second period self will not work no matter what.

In period 1, the agent chooses not to work. Starting a career is fruitless because her future selves would not stay in the labor force. The first period self cannot commit her future selves to work, and for this reason decides not to work herself. If, however, she could commit her future selves to work, she would have preferred a life of work to a life on welfare because

$$w(0) - c_h + \beta [w(1) - c_l + w(2) - c_l] = 2.25 > b + \frac{1}{2}[b + b] = 2.$$

Next we consider the agent's attitude toward a one-period time limit, and how that attitude changes depending on when she is asked.

First Attitude

If asked at the beginning of period 1 (or at any time prior to the decision making modeled here) whether she would support the implementation of a one-period time limit, the agent would answer "Yes, I would like such a limit." She prefers the time limit because it commits her future selves to work. Consider the agent in period two. If the agent has worked in period 1, i.e., if $\tau_2 = 1, d_1 = 1$, the period 2 self's payoff from working would be

$$w(1) - c_l + \beta [w(2) - c_l] = -9 - 1 + \frac{1}{2}[23.5 - 1] = 1.25.$$

If instead she stays on welfare in period 2, the third period self would stay at home (because of the time limit), thus her expected payoff from participating in welfare in period 2 is given by:

$$1 + \beta [0] = 1.$$

The second period self works if the period one self worked. Anticipating this, the period one self would work and, as we showed earlier, achieve a higher discounted payoff than that from a life on welfare. She thus strictly prefers that the time limit be implemented.[19]

Later Attitudes

After the policy is implemented, however, the agent's attitude changes. If asked then whether she would prefer the time limit remain in place, she would either say "No, I prefer that the limit be relaxed," or she would say she is indifferent. In period 2, having worked in the previous period she would strictly prefer to relax the welfare time limit, choose welfare today and anticipate a payoff of:

$$1 + \beta [1] = 1.5$$

which strictly exceeds her payoff under the time limit. In period 3 if we ask the agent the same question she will say she is indifferent, she will work no matter what.

C.2 Attitudes of a Projection-biased Agent Toward Time Limit

Now we analyze a projection-biased agent's preferences for policy. We again first consider behavior in the absence of a time limit. In period three, optimal choice is like that of a present biased agent; the agent works only if she worked in the previous two periods. In period 2, if $\tau_2 = 1$, then $d_1 = 1$ and the payoff from working is given by

$$w(1) - c_l + w(2) - c_l = -9 - 1 + 23.5 - 1 > 2$$

so she works. If $\tau_2 = 0$, then $d_1 = 0$, and the expected payoff from working is $w(0) - c_h + b < 2b$ so she will not work. Thus, in period 2 the agent works only if she worked the previous period.

In period 1, the agent foresees that if she chooses work, she will work for the rest of her career. However, projection bias leads her to perceive the utility from working as

$$\begin{aligned} & [w(0) - c_h] + [w(1) - c_h] + [w(2) - c_h] \\ = \;& 0 - 4 + [-9 - 4] + [23.5 - 4] = 2.5 < 3b = 3. \end{aligned}$$

Thus in period 1, the agent chooses welfare because she over-predicts the future cost of work.

First Attitude

[19] An unbiased agent would, of course, never strictly prefer a time limit.

Now consider the agent's attitudes toward a one-period time limit. In period one (or at any time she is unemployed prior to the decision making modeled here) she opposes the idea. She opposes the limit her choices are now choose work and anticipate payoffs

$$[w(0) - c_h] + [w(1) - c_h] + [w(2) - c_h] = 2.5$$

or choose welfare and anticipate

$$b + 0 + 0 = 1.$$

Each is strictly dominated, from her perspective, by a lifetime on welfare.

Later Attitudes

Sometime later, however, after the policy has been implemented, the agent's attitude changes. If asked then whether she would prefer that the time limit remain in place, she would say "I am indifferent." In period 2 or 3, having worked in the previous period she would strictly prefer to work regardless of the limit on welfare. If the horizon were longer and, more realistically, $\rho_h, \rho_l < 1$ then a working agent could strictly prefer that the time limit remain in place. In this case, if the agent fears losing her job, she would properly be concerned that once unemployed she would come to view the costs of working as too high and remain stuck on welfare.

References

[1] Ainslie, G., and N. Haslam. 1992. "Hyperbolic Discounting," in G. Loewenstein and J. Elster, eds., *Choice over Time*. New York: Russell Sage Foundation.

[2] Arkin, R.M. and G.M. Maruyama. 1979. "Attribution, Affect and College Exam Performance." *Journal of Educational Psychology*, **71**: 85-93.

[3] Aronson, E. 1994. *The Social Animal.* 7th Ed., W.H. Freeman and Company: NewYork.

[4] Ameriks, John, Andrew Caplin, and John Leahy, 2003, "Wealth Accumulation and the Propensity to Plan." *Quarterly Journal of Economics* **118(3)**: 1007-1047.

[5] Babcock, L., C. Camerer, G. Loewenstein, and R. Thaler. 1997. "Labor Supply of New York City Cab Drivers: One Day at a Time." *Quarterly Journal of Economics,* **111**: 408-441.

[6] Barberis, Nicholas and Ming Huang. 2001. "Mental Accounting, Loss Aversion, and Individual Stock Returns." *Journal of Finance,* **56(4)**: 1247-1292.

[7] Barro, Robert. 1999. "Ramsey Meets Laibson in the Neoclassical Growth Model." *Quarterly Journal of Economics,* **114(4)**: 1125-1152.

[8] Barsky, Robert, F. Thomas Juster, Miles Kimball, and Mathew Shapiro, 1997, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, **112(2)**:537-579.

[9] Cunningham, J.D., P.A. Starr and D.E. Kanouse. 1979. "Self as Actor, Active Observer, and Passive Observer: Implications for Causal Attribution." *Journal of Personality and Social Psychology*, **37**: 1146-1152.

[10] Della Vigna, Stefano, and Ulrike Malmendier. 2002. "Overestimating Self-Control: Evidence from the Health Club Industry." Mimeo, University of California.

[11] Benartzi, Shlomo and Richard Thaler, 2004, "Save More Tomorrow^{TM} :Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, **112(S1):**S164-S187.

[12] Bertrand, Marianne and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review Papers and Proceedings*, **91(2)**, 67-72.

[13] Edin, Kathryn and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survive Welfare and Low-wage Work*. New York: Russell Sage.

[14] Fang, Hanming and Dan Silverman. 2004a. "On the Compassion of Time-Limited Welfare Programs." *Journal of Public Economics,* **88(7-8)**: 1445-1470.

[15] ———— and ————. 2004b. "Time-inconsistency and Welfare Program Participation: Evidence from the NLSY." Cowles Foundation Discussion Paper 1465, Yale University.

[16] Frederick, Shane, George Loewenstein and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, **XL**, 351-401.

[17] Genesove, David and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics,* **116(4)**: 1233-1260.

[18] Gilbert, D. T., E. C. Pinel, T. D. Wilson, S. J. Blumberg, and T. P. Wheatley, 1998. "Immune Neglect: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology,* **75(3)**:617-638.

[19] Gilovich, T. 1983. "Biased Evaluation and Persistence in Gambling." *Journal of Personality and Social Psychology*, **44**: 1110-1126.

[20] Gruber, Jonathan, and Botond Koszegi, 2001. "Is Addiction 'Rational?' Theory and Evidence." *Quarterly Journal of Economics*, **116(4)**: 1261-1305.

[21] Gruber, Jonathan and Sendhil Mullainathan. 2002. "Do Cigarette Taxes Make Smokers Happier?" mimeo, MIT.

[22] Harris, Christopher and David Laibson. 2001. "Dynamic Choices of Hyperbolic Consumers." *Econometrica,* **69(3)**: 397-421.

[23] Hotz, Joseph and Robert Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies,* **60(3)**: 497-530.

[24] Johnston, W.A. 1967. "Individual Performance and Self-Evaluation in a Simulated Team." *Organization Behavior and Human Performance*, **2**: 309-328.

[25] Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics,* **112:** 443-477.

[26] Laibson, David, Andrea Repetto, Jeremy Tobacman. 1998. "Self-Control and Saving for Retirement." *Brookings Papers on Economic Activity,* **1:** 91-196.

[27] Larwood, L. and W. Whittaker. 1977. "Managerial Myopia: Self-serving Biases in Organizational Planning." *Journal of Applied Psychology*, **62**: 194-198.

[28] Loewenstein, George, and Jon Elster. 1992. *Choice Over Time.* Russell Sage: New York.

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

[30] Loewenstein, George and David Schkade. 1999. "Wouldn't It Be Nice? Predicting Future Feelings," in Daniel Kahneman, Edward Diener and Norbert Schwarz eds., *Well-Being: The Foundations of Hedonic Psychology* (New York, NY: Russell Sage Foundation Press).

[31] Manski, Charles F. 2004. "Measuring Expectations." Forthcoming, *Econometrica*.

[32] O'Donoghue, Ted and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review,* **89(1)**: 103-124.

[33] O'Donoghue, Ted and Matthew Rabin. 2001. "Choice and Procrastination." *Quarterly Journal of Economics,* **166(1)**: 121-160.

[34] Paserman, M. Daniele. 2002. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation." Mimeo, Hebrew University.

[35] Phelps, Edmund S., and Robert A. Pollak. 1968. "On Second-best National Saving and Game-equilibrium Growth." *Review of Economic Studies,* **35(2)**: 185-199.

[36] Pollak, Robert A. 1968. "Consistent Planning." *Review of Economic Studies,* **35(2)**: 201-208.

[37] Rabin, Matthew, and Joel Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics,* **114(1)**:37-82.

[38] Read, Daniel and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* **76(2)**: 189-205.

[39] Riis, Jason, George Loewenstein, Jonathan Baron, Christopher Jepson, Angela Fagerlin and Peter Ubel, 2004. "Ignorance of Hedonic Adaptation to Hemo-Dialysis: A Study Using Ecological Momentary Assessment," in press *Journal of Experimental Psychology: General.*

[40] Ross, M. and F. Sicoly. 1979. "Egocentric Biases in Availability and Attribution." *Journal of Personality and Social Psychology*, **37**: 322-336.

[41] Rubinstein, Ariel. 2003. "Economics and Psychology"? The Case of Hyperbolic Discounting. *International Economic Review,* **44**, 1207-1216.

[42] Sackett, D. J., and G.W. Torrance. 1978. "The Utility of Different Health States and Perceived by the General Public." *Journal of Chronic Diseases,* 31:697-704.

[43] Shapiro, Matthew and Joel Slemrod. 2003. "Consumer Response to Tax Rebates." *American Economic Review*, **93(1)**: 381-396.

[44] Strotz, Robert H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies,* **23(3):** 165-180.

[45] Svenson, O. 1981. "Are We All Less Risky and More Skillful than Our Fellow Drivers?" *Acta Psychologica*, **47**: 143-148.

[46] Wertheimer, Richard, Melissa Long, and Sharon Vandivere. 2001. "Welfare Recipients' Attitudes toward Welfare, Nonmarital Childbearing, and Work: Implications for Reform?" New Federalism: National Survey of America's Families Report #310300, Urban Institute Press, Washington D.C.