CAE Working Paper #02-11

Projection Bias in Predicting Future Utility

by

George Lowenstein
Ted O'Donoghue
and
Matthew Rabin

August 2002

.

Projection Bias in Predicting Future Utility

George Loewenstein
Carnegie-Mellon University

Ted O'Donoghue
Cornell University

Matthew Rabin
University of California - Berkeley

July 24, 2002

Abstract

People exaggerate the degree to which their future tastes will resemble their current tastes. We present evidence from a variety of domains which demonstrates the prevalence of such *projection bias*, develop a formal model of it, and use this model to demonstrate its importance in economic environments. Projection bias leads people to overvalue reference-dependent goods and magnifies the endowment effect. It can cause misguided purchases of durable goods. When there is habit formation, projection bias can lead people to consume too much early in life, and to decide, as time passes, to consume more — and save less — than originally planned.

Keywords: Changing Tastes, Dynamic Inconsistency, Impulse Purchases, Intertemporal Choice, Misprediction, Reference Dependence.

JEL Classification: A12, B49, D11, D91, E21

Acknowledgments: For helpful comments, we are grateful to Erik Eyster, Chris Harris, and members of the Russell Sage Foundation Behavioral Economics Roundtable; seminar participants at Cornell University, Yale University, Harvard University, University of Michigan, University of Texas, London School of Economics, University of Zurich, the Toulouse Conference on Psychology and Economics, and the Jerome Levy Institute; Larry Katz and an anonymous referee; and especially Colin Camerer and Drazen Prelec for very helpful discussions at the formative stages of this project. For research assistance, we thank Kitt Carpenter, Erik Eyster, Jeff Holman, David Huffman, Chris Meissner, and Mandar Oak. For financial support, Loewenstein thanks the Center for the Study of Human Dimensions of Global Change at Carnegie Mellon University (NSF Grant SBR-9521914), O'Donoghue and Rabin thank the National Science Foundation (Awards SBR-9709485, SES-0078796, and SES-0079266), and Rabin thanks the Russell Sage, MacArthur, and Sloan Foundations. This research was started while Loewenstein and Rabin were Fellows at the Center for Advanced Study in the Behavioral Sciences, supported by NSF Grant SBR-960123, and they are very grateful for the Center's hospitality and the NSF's support.

Mail: George Loewenstein / Department of Social and Decision Sciences / Carnegie Mellon University /Pittsburgh, PA 15213-3890; Ted O'Donoghue / Department of Economics / Cornell University / 414 Uris Hall / Ithaca, NY 14853-7601; Matthew Rabin / Department of Economics / 549 Evans Hall #3880 / University of California, Berkeley / Berkeley, CA 94720-3880. E-mail: gl20@andrew.cmu.edu, edo1@cornell.edu, and rabin@econ.berkeley.edu. Web pages: sds.hss.cmu.edu/faculty/loewenstein.html, www.people.cornell.edu/pages/edo1/, and emlab.berkeley.edu/users/rabin/.

"The great source of both the misery and disorders of human life, seems to arise from the over-rating the difference between one permanent situation and another. Avarice over-rates the difference between poverty and riches: ambition, that between a private and public station: vain-glory, that between obscurity and extensive reputation."

— Adam Smith, *The Theory of Moral Sentiments* (2002; p. 173; III,iii,31).

1. Introduction

A person's tastes typically change over time, due to such factors as past consumption, day-to-day mood fluctuations, social influence, maturation, and changes in the environment. Optimal decision making often requires a person to predict and take account of future changes in tastes. When making summer vacation plans during the winter, for instance, a person must predict how she will feel in the summer. When deciding whether to try cigarettes, she must predict how this consumption will influence her future enjoyment of activities, including smoking further cigarettes.

In this paper, we provide evidence for, formalize, and explore the implications of a general bias in such predictions: People tend to exaggerate the degree to which their future tastes will resemble their current tastes. We investigate the nature of errors that this *projection bias* can lead to in dynamic-choice environments.

In Section 2, we review evidence from a variety of domains supporting the existence of projection bias. In environments in which tastes change over time, people tend to understand qualitatively the directions in which their tastes will change, but underestimate the magnitudes of these changes. People underappreciate the effects of frequently fluctuating tastes, such as fluctuating hunger. Perhaps more importantly, people also underappreciate the effects of longer-term changes, such as those that result from adaptation to a shifting standard of living. Indeed, virtually all evidence we are familiar with on misprediction of future tastes is consistent with projection bias.

In Section 3, we develop a formal model of projection bias. Projection bias influences a person's predictions of her future tastes. To fix ideas, suppose a person's instantaneous utility can be written as $u(c, s)$, where c is her consumption and s is a "state" that parameterizes her tastes. Suppose further that the person with current state s' must predict her tastes at a time in the future when her state will be s. The evidence suggests that the person's prediction, $\widetilde{u}(c, s|s')$, typically lies somewhere "in between" her true future tastes $u(c, s)$ and her current tastes $u(c, s')$. While we provide a more general definition in Appendix A, our formal analysis in this paper assumes $\widetilde{u}(c, s|s')$ is a simple linear combination of $u(c, s)$ and $u(c, s')$, which we refer to as *simple projection bias.*

Because it implies that predicted utilities need not match actual utilities, projection bias predicts that a person's behavior need not correspond to correct intertemporal utility maximization. If, for instance, current consumption has deleterious effects on future well-being, and projection bias leads the person to underappreciate these effects, she may over-consume relative to what would maximize her true intertemporal utility. Moreover, as a person's tastes change over time in ways she did not predict, she may not stick to earlier plans — that is, projection bias can give rise to dynamic inconsistency. A stressed undergraduate who underappreciates the addictiveness of cigarettes, for instance, might start smoking with the expectation of quitting upon graduation, only to continue smoking after graduation due to the unanticipated addiction.

In Sections 4, 5, and 6, we illustrate the potential importance of projection bias for economics by analyzing its implications in three different environments. Section 4 describes its implications for the experimentally-established endowment effect, whereby people tend to value objects more highly if they possess them than if they do not. We present a simple model that incorporates the usual explanation of the endowment effect: People have reference-dependent preferences (they experience feelings of gain or loss upon obtaining or parting with objects), and are loss-averse (they experience losses more intensely than gains). We first show that projection bias creates a tendency to over-value goods, because people exaggerate the degree to which feelings of gain will persist when buying goods and the degree to which feelings of loss will persist when selling goods. More interesting, projection bias magnifies the size of the endowment effect — that is, if people have projection bias, they will exhibit an endowment effect that is larger than their own subsequent feelings of loss and gain justify. Hence, while the endowment effect may be caused by a valid expectation that losses will hurt more than gains will help, projection bias leads people to project this difference further into the future than is justified. Finally, we explore the implications of projection bias in a (second-hand) market setting, and conclude that projection bias leads to diminished volume of trade and increased market prices.

In Section 5, we show how projection bias can cause misguided purchases of durable goods. The satisfaction that a person derives from a durable good might change over time for at least two reasons. First, her valuation is likely to systematically decline over time as the "novelty" of the item wears off. Second, there are likely to be random, day-to-day fluctuations in her valuation. While projection bias over decay in tastes unambiguously leads a person to over-value the durable good, projection bias over day-to-day fluctuations variously leads a person to over-value or under-value the good. When a person projects an above-average current valuation onto her prediction of future valuations, she will over-value the good; when she projects a below-average current valuation onto her prediction of future valuations, she will under-value the good. Hence, a person making a one-time buying decision is equally likely to buy when she shouldn't or not to buy when she should. If a person has multiple opportunities to buy, however, an inherent asymmetry in the purchasing decision means projection bias will lead a person to more often buy durable goods when she shouldn't. Since un-buying is typically much harder than buying, a person will buy a durable good when she shouldn't if on *any* day she incorrectly thinks it is worth it, while she won't buy it when she should only if she *always* incorrectly thinks it is not worth it.

In Section 6, we explore the implications of projection bias in a life-cycle consumption model with habit formation. Most models of habit formation incorporate two features: Current consumption increases future *marginal* utility (which is more or less what is meant by habit formation) but reduces the future *level* of utility. We show that, because projection bias leads a person to underappreciate these forces, she chooses a consumption profile that involves too much consumption early in life and too little consumption late in life relative to what would be optimal. Moreover, as time passes and the person habituates to higher consumption levels, she may decide to consume more than she had earlier planned; hence projection bias can cause saving to fall short of intentions. Finally, as the person gets accustomed to higher consumption levels, she also values income more highly, and hence might decide to work more (or retire later) than she had earlier planned.

We conclude in Section 7 by putting projection bias in broader economic context, and discussing some shortcomings and potential extensions of our model.

2. Evidence of Projection Bias

In this section, we review evidence from a variety of domains that supports our contention that, while people understand qualitatively the many ways in which their tastes might change over time, they tend to systematically underestimate the magnitudes of these changes.[1]

We begin with the domain of adaptation. A plethora of evidence demonstrates that adaptation is a central component of the human experience (see Helson (1964), and Frederick and Loewenstein (1999) for a recent review). People have a remarkable ability to adapt to major changes in their life circumstances, such as incarceration, health conditions, moving to a different climate or occupation, and so forth.[2] When changes in tastes are driven by adaptation, projection bias implies that people

[1] See Loewenstein and Schkade (1999) for a summary of much of the evidence presented in this section, as well as for a discussion of the psychological mechanisms that underlie projection bias.
[2] There are some exceptions to this rule. First, there are a variety of factors that impede adap-

underappreciate the extent to which they will adapt to new circumstances, and hence overestimate the impact of major changes on their long-run level of happiness. A great deal of evidence supports this prediction.

Suggestive evidence comes from happiness comparisons across groups. In a classic study, Brickman, Coates, and Janoff-Bulman (1978) interviewed people who had won lottery jackpot prizes within the last year (average winnings of $479,545), people with paraplegia, and a control group. They found virtually no difference in self-reported happiness between lottery winners and the control group, and a surprisingly small difference in self-reported happiness between paraplegics and the control group (2.96 for the paraplegics vs. 3.82 for the control group on a 5 point scale). Other researchers have since found similar results. Schulz and Decker (1985) found that reported well-being levels of elderly paraplegics and quadriplegics were only slightly lower than population means of non-disabled people of similar age. Wortman and Silver (1987) found that quadriplegics reported no greater frequency of negative affect than control respondents. Tyc (1992) found "no difference in quality of life or psychiatric symptomatology" in young patients who had lost limbs to cancer compared with those who had not. While these papers present no data on how people *predict* they would feel if they won the lottery or became disabled, the notion that lottery winners are no happier and that paraplegics are only slightly less happy certainly runs counter to the predictions of most people — including, presumably, those playing the lottery.

Better evidence comes from studies that compare one group's predictions with another group's self-reports. Several studies have found that non-patients' predictions of the quality of life associated with serious medical conditions are lower than actual patients' self-reported quality of life. For example, on a 0 to 1 scale on which 0 means as bad as death and 1 means perfect health, non-patients predict that chronic dialysis would yield a quality of life of 0.39, whereas dialysis patients say it yields a quality of life of 0.56 (Sackett and Torrance, 1978); patients with colostomies rate their own quality of life at .92, while non-patients predict that if they had a colostomy, they would rate their quality of life at 0.80 (Boyd, Sutherland, Heasman, Tritcher, and Cummings, 1990). There is also longitudinal evidence in the medical domain. Jepson, Loewenstein, and Ubel (2001) asked people waiting for a kidney transplant to predict what their quality of life would be one year later if they did or did not receive a transplant, and then asked those same people one year later to

tation, such as uncertainty about whether a situation is permanent and repeated reminders of the original situation. Second, some studies have found that people do not seem to adapt to noise; indeed, if anything, they seem to become increasingly irritated by it (for an overview, see Weinstein, 1982). Moreover, noise is the one example we know of that might contradict our assertion that people understand the direction in which tastes change, because people seem predict that they will adapt when in fact they tend to become more irritated.

report their current quality of life. Echoing the pattern found in cross-sectional studies, patients who received transplants predicted that they would feel better than they ended up feeling, and those who did not expected to feel worse than they actually did. Sieff, Dawes, and Loewenstein (1999) found similar longitudinal results for people being tested for HIV.

Evidence on underappreciation of adaptation is not limited to the medical domain. Gilbert, Pinel, Wilson, Blumberg, and Wheatley (1998) reported several instances of people underestimating adaptation to unfavorable events — which they label "immune neglect." For example, assistant professors at the University of Texas were asked to forecast their subjective well-being at various points in time following their tenure decision, conditional on the decision being favorable and unfavorable. At the same time, people who had been assistant professors at the University of Texas during the previous ten years, and received either a positive or negative tenure decision, were asked to report their subjective well-being. A comparison of these groups revealed that subjects exaggerated the longevity of the impact of tenure: People were relatively accurate in predicting the immediate impact of getting or being denied tenure, but they extrapolated these positive and negative feelings further into the future than warranted. In a similar study, Loewenstein and Frederick (1997) compared the predictions by survey respondents of how various events would affect their well-being over the next decade to the reports of other respondents about how actual events in the past decade had affected their well-being. The events included environmental changes (e.g., decline in sport-fishing), social changes (e.g., increase in number of coffee shops), and personal changes (e.g., increases in body weight or income). A clear pattern emerged in the data: Those making prospective predictions expected future changes to affect their well-being more than those making retrospective evaluations reported that matched changes in the past had affected their well-being.

An important methodological issue arises for interpreting all of the studies above: Do the similarities in hedonic ratings reflect true comparability in hedonic experiences, or do they rather reflect that responses are scaled differently across groups? Paraplegics, for instance, might interpret a 0.8 on a 0-to-1 scale differently from control subjects because they implicitly rate their own happiness relative either to other paraplegics or to the extreme despair they experienced immediately following the onset of their disability. Both of these tendencies would result in paraplegics giving misleadingly high ratings of their own quality of life, which could explain the similarities in reported happiness between paraplegics and non-paraplegics, as well as the discrepancy between patients' ratings and non-patients' predictions, even if actual experiences were very different and correctly predicted.

We believe this problem is a pervasive and important caveat to all of the studies above. But we also believe that some results militate against interpreting the evidence above as solely a matter of such "response norming". First, there is evidence that is it possible to "debias" people to some extent — to bring nonpatients' predictions of the quality of life closer to patients' ratings — by inducing people to think more carefully about adaptation (Ubel, Loewenstein, Jepson, Mohr, and Markowitz, 2002). One would not observe such effects if the discrepancy between patients and nonpatients resulted only from differences in the way they interpret the quality-of-life scale. Second, there is evidence that people who are personally familiar with paraplegics and lottery winners give quality-of-life responses that are more in line with the responses of those who have experienced such outcomes, which is predicted by projection bias, but not by scale norming. Cohn (1999; cited in Kahneman, 2000) asked subjects to estimate the percentage of time that paraplegics and lottery winners were in a good, neutral, or bad mood either one month or one year after the event. Subjects who did not know a lottery winner or paraplegic predicted very little adaptation — that is, their predictions for one month vs. one year after the event were roughly the same. Subjects who knew a lottery winner or paraplegic, in contrast, predicted substantial adaptation. Those who were personally familiar with one of these groups, it seems, had gleaned from their observations a greater appreciation for the power of adaptation. Finally, our confidence in the projection-bias account of these findings is bolstered by the similarity of these results with those from controlled laboratory settings focusing on predictions of short-term changes in tastes. We turn now to a discussion of this evidence.

We begin with experiments that document an underappreciation of the *endowment effect*, which, as introduced by Thaler (1980), refers to people's tendency to value an object more highly if they possess it than if they do not. Endowment-effect experiments demonstrate that the reservation prices of *sellers* — subjects endowed with the object who have the option to sell — are typically larger than the reservation prices of both *buyers* — subjects not endowed with the object who have the option to buy — and *choosers* — subjects not endowed with the object who choose between the object and some amount of money.[3] When tastes change according to the endowment effect, projection bias implies that unendowed subjects will underestimate by how much becoming endowed will increase their valuation, and that endowed subjects will underestimate by how much becoming unendowed will decrease their valuation.

While we know of no direct evidence on the latter implication, Loewenstein and Adler (1995)

---

[3] See Kahneman, Knetsch, and Thaler (1991) for a review of the endowment effect. Comparing sellers and choosers is often preferred because it allows one to ignore both the legitimate concern about the role of loss aversion over money and the rather silly (but sometimes raised) concern about wealth effects.

find evidence of the former. In one study, subjects randomly assigned to a "prediction" treatment group were shown an embossed coffee mug and told to imagine that they were given one as a prize and had the opportunity to exchange it for cash. They were then shown the form that would be used to elicit their selling price and were asked to complete it as they expected they would once they received the mug. After a delay, they were actually given the mug, and then asked to complete the same form eliciting selling prices. The other half of subjects were simply given mugs without first making predictions, and then they completed the form eliciting selling prices. The results, presented in Table 1, reveal a systematic underappreciation of the impact of endowment on preferences: The predicted selling prices of the prediction group were substantially lower than the actual selling prices of both the prediction group and the non-prediction group.

Table 1: Predicted and Actual Valuation of Mug
(from Loewenstein and Adler (1995))

Group	Condition	Number of Subjects	Prediction of Valuation	Actual Valuation
Carnegie Mellon University	Prediction	14	\$3.73 (0.41)	\$5.40 (0.65)
	No Prediction	13	———	\$6.46 (0.54)
University of Pittsburgh	Prediction	22	\$3.27 (0.48)	\$4.56 (0.59)
	No Prediction	17	———	\$4.98 (0.53)

(standard errors in parentheses)

Indirect evidence of underappreciation of the endowment effect comes from a series of studies by Van Boven, Dunning, and Loewenstein (2000) on interpersonal predictions. In one experiment, the usual endowment effect was replicated by eliciting selling prices from subjects endowed with coffee mugs and buying prices from subjects not endowed. Sellers were then asked to estimate how much buyers would pay, and buyers were asked to estimate how much sellers would charge, with all subjects rewarded for accurate predictions. Subjects exhibited an "interpersonal projection bias": Sellers over-estimated buying prices, and buyers under-estimated selling prices. Follow-up experiments support the conclusion that this interpersonal projection bias results from intrapersonal projection bias — from the subjects' tendency to mispredict the price at which they themselves would buy or sell the mug.

The final domain for which we review specific evidence is underappreciation of the effects of hunger. This evidence is particularly valuable because fluctuations in hunger are something that people ought to understand well. If people exhibit projection bias with regard to fluctuations in hunger, then we shouldn't be surprised if they exhibit the same systematic tendency in other realms where they have less opportunity to learn.

Several studies lend support to the folk wisdom that shopping on an empty stomach leads people to buy too much (Nisbett and Kanouse (1968), Gilbert, Gill, and Wilson (1998)). This phenomenon can be interpreted as a manifestation of projection bias: People who are hungry act as if their future taste for food will reflect such hunger. Read and van Leeuwen (1998) provide even sharper evidence of projection bias with respect to hunger. Office workers were asked to choose between healthy snacks and unhealthy snacks that they would receive in one week, either at a time when they should expect to be hungry (late in the afternoon) or satiated (immediately after lunch).[4] Subjects were approached to make the choice either when they were hungry (late in the afternoon) or satiated (immediately after lunch). As depicted in Table 2, people who expected to be hungry the next week were more likely to opt for unhealthy snacks than those who expected to be satiated, presumably reflecting an increased taste for unhealthy snacks in the hungry state. But in addition, people who were hungry *when they made the choice* were more likely to opt for unhealthy snacks than those who were satiated, suggesting that people were projecting their current tastes onto their future tastes.

Table 2: Percentage of Subjects Choosing Unhealthy Snack
(from Read and van Leeuwen (1998))

		Future Hunger	
		Hungry	Satiated
Current Hunger	Hungry	78%	56%
	Satiated	42%	26%

Indeed, if we interpret the main diagonal — the hungry-hungry condition and the satiated-satiated condition — as reflecting true preferences, then the data fits exactly the pattern of projection bias. For instance, for those subjects who are currently satiated but expect to be hungry, they

[4] The healthy snacks were apples and bananas; the unhealthy snacks were crisps, borrelnoten, Mars Bars, and Snickers Bars. We adopt the terminology healthy and unhealthy from the experimenters.

understand the direction in which their tastes will change as they become hungry — more choose the unhealthy snack than in the satiated-satiated condition — but they underestimate the magnitude of this change — fewer choose the unhealthy snack than in the hungry-hungry condition. An analogous conclusion holds for subjects who are currently hungry and expect to be satiated.

While we have limited our detailed discussion to a few studies, there is considerable further evidence that projection bias is a phenomena that operates in a consistent fashion across a broad array of domains. Indeed, virtually all evidence that we are aware of is consistent with projection bias (except possibly noise, as discussed in Footnote 2).[5] Projection bias also resembles other psychological phenomena involving judgment rather than choice. Just as projection bias is characterized by predictions of future tastes that lie between true future tastes and current tastes, there are a number of judgmental biases in which people's judgments lie between the truth and some naive benchmark. Three examples are the "hindsight bias" (Fischhoff, 1975) — after an event occurs, people overestimate the degree to which they could have predicted the event before it occurred — the "false-consensus effect" (Ross, Greene, and House, 1977) — the tendency to overestimate the degree to which others will think, feel, or behave similarly to oneself — and the "curse of knowledge" (Camerer, Loewenstein, and Weber, 1992) — the tendency to think that the information that one possesses is more widely shared than it actually is. Indeed, the formalization of the curse of knowledge used by Camerer, Loewenstein, and Weber (1992) is strikingly parallel to our formalization of simple projection bias in the next section. The resemblance of projection bias to these other, more cognitive, phenomena raises the question of whether they share one or more mechanisms. A plausible contender for such a mechanism is the more basic psychological notion of anchoring and adjustment. With regard to projection bias, this would mean that when making predictions about future tastes, people anchor on their current tastes and then insufficiently adjust to account for changes in tastes.

Based on our belief that projection bias describes patterns of behavior in a broad array of environments, our goal in the remainder of this paper is to demonstrate its potential importance in economic environments. We do so by building a formal model of projection bias and then analyzing its implications in some specific economic environments.

---

[5] Other domains for which there is evidence consistent with projection bias include sexual arousal (Loewenstein, Nagin, and Paternoster, 1997), pain (Read and Loewenstein, 1999), thirst (Van Boven and Loewenstein, forthcoming), and heroin craving (Giordano et al., 2001). See also Loewenstein's (1996,1999) discussion of hot/cold empathy gaps wherein individuals who are in cold visceral states underappreciate the impact of hot visceral states on their own behavior.

3. The Model

Projection bias makes predictions both about behavior and about the nature of errors in utility maximization, but only once we know how true tastes change. It will never provide a prediction without first designating preferences for the particular situation. Hence, our starting point, both in our general abstract model and in our specific applications, is to specify the ways in which true tastes change.[6]

Suppose a person's true intertemporal preferences are given by

$$U^t = \sum_{\tau=t}^{T} \delta^\tau u(\mathbf{c}_\tau, \mathbf{s}_\tau),$$

where $u(\mathbf{c}_\tau, \mathbf{s}_\tau)$ is her instantaneous utility in period τ, $\delta \leq 1$ is her discount factor, and T is her (possibly infinite) time horizon. The vector $\mathbf{c}_\tau$ is the person's period-τ consumption vector; $\mathbf{c}_\tau$ includes all period-τ behavior relevant for current or future instantaneous utilities. The vector $\mathbf{s}_\tau$ is the person's "state" in period τ, which parametrizes her tastes in period τ. As such, the state incorporates all factors that affect instantaneous utility besides current consumption, including past consumption — as when past consumption of a good determines current addiction to that good — and exogenous factors — as when fluctuations in serotonin levels affect mood or when peer pressure affects the benefits and costs of current behavior.[7]

Models with such state-dependent preferences are becoming more common in economics, particularly in the realm of habit formation that we analyze in Section 6. When studying such preferences, economists typically assume that the person is "fully rational" in the sense that she correctly predicts how current consumption affects the evolution of future states and hence how current consumption affects all future utilities. Formally, for any period t and initial state $\mathbf{s}_t$, a fully rational person chooses a path of consumption $(\mathbf{c}_t, ..., \mathbf{c}_T)$, correctly anticipating the associated path of states $(\mathbf{s}_t, ..., \mathbf{s}_T)$, to maximize true intertemporal utility U^t.

In our model, a person *attempts* to maximize her intertemporal utility, but she may fail to do so because she mispredicts her future instantaneous utilities. We assume that the person understands the evolution of future states — for any consumption plan $(\mathbf{c}_t, ..., \mathbf{c}_T)$ she correctly anticipates the associated path of states $(\mathbf{s}_t, ..., \mathbf{s}_T)$ — but she mispredicts the impact of future states on her future utility.[8] Let $\widetilde{u}(\mathbf{c}, \mathbf{s}|\mathbf{s}')$ denote the prediction of a person currently in state $\mathbf{s}'$ for what her future

6 For an overview of taste changes, see Loewenstein and Angner (forthcoming).

7 While we assume throughout the paper that the utility function itself is not a function of the date, the model could be extended by treating calendar time as a state variable.

8 Our model is essentially equivalent to an alternative formulation wherein people underestimate the degree to which the states will change.

instantaneous utility would be from consuming $\mathbf{c}$ in state $\mathbf{s}$. For a fully rational person, predicted utility equals true utility — that is, $\widetilde{u}(\mathbf{c},\mathbf{s}|\mathbf{s}') = u(\mathbf{c},\mathbf{s})$. A person with *projection bias* understands the qualitative nature of changes in her preferences, but underestimates the magnitude of these changes. Roughly speaking, her predicted utility $\widetilde{u}(\cdot,\mathbf{s}|\mathbf{s}')$ lies "in between" her true utility function $u(\cdot,\mathbf{s})$ and utility in the current state $u(\cdot,\mathbf{s}')$. In this paper, we consider a particularly simple form of projection bias:

Definition 1. Predicted utility exhibits *simple projection bias* if there exists $\alpha \in [0,1]$ such that for all $\mathbf{c}$, $\mathbf{s}$, and $\mathbf{s}'$, $\widetilde{u}(\mathbf{c},\mathbf{s}|\mathbf{s}') = (1-\alpha)\ u(\mathbf{c},\mathbf{s}) + \alpha\ u(\mathbf{c},\mathbf{s}')$.

With this formulation, if $\alpha = 0$, the person has no projection bias: She predicts her future instantaneous utility correctly. If $\alpha > 0$, the person has projection bias; the bigger is α, the stronger is the bias. When $\alpha = 1$, the person perceives that her future tastes will be identical to her current tastes.[9]

If a person exhibits projection bias and her state in period t is $\mathbf{s}_t$, then she perceives her period-t intertemporal utility for consumption profile $(\mathbf{c}_t, ..., \mathbf{c}_T)$ to be

$$\widetilde{U}^t = \sum_{\tau=t}^{T} \delta^\tau \widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}_t).$$

We assume that for any period t and initial state $\mathbf{s}_t$, a person with projection bias chooses a path of consumption $(\mathbf{c}_t, ..., \mathbf{c}_T)$, correctly anticipating the associated path of states $(\mathbf{s}_t, ..., \mathbf{s}_T)$, to maximize her perceived intertemporal utility $\widetilde{U}^t$. That is, she behaves exactly as a fully rational person would except that $\widetilde{U}^t \neq U^t$.

To incorporate uncertainty over future consumption or future states, we make the standard assumption that a person maximizes her expected discounted utility. For instance, suppose that in period t the person expects her period-τ consumption-state combination to be $(\mathbf{c}', \mathbf{s}')$ with probability p and $(\mathbf{c}'', \mathbf{s}'')$ with probability $1-p$. Just as true period-τ expected utility is $E_t\left[u(\mathbf{c}_\tau, \mathbf{s}_\tau)\right] = p\ u(\mathbf{c}', \mathbf{s}') + (1-p)u(\mathbf{c}'', \mathbf{s}'')$, a person with projection bias predicts period-τ expected utility to be $E_t\left[\widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}_t)\right] = p\ \widetilde{u}(\mathbf{c}', \mathbf{s}'|\mathbf{s}_t) + (1-p)\widetilde{u}(\mathbf{c}'', \mathbf{s}''|\mathbf{s}_t)$. Similarly, true expected intertemporal utility is $E_t\left[U^t\right] = E_t\left[\sum_{\tau=t}^{T} \delta^\tau u(\mathbf{c}_\tau, \mathbf{s}_\tau)\right]$, and a person with projection bias perceives her expected

[9] While simple projection bias is sufficient for our analysis in this paper, it is too restrictive for use as a general definition. In Appendix A, we describe some of the limitations of this definition, and provide a more general formulation of projection bias.

intertemporal utility to be $E_t\left[\widetilde{U}^t\right] = E_t\left[\sum_{\tau=t}^{T}\delta^\tau \widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|s_t)\right]$.[10]

In our model, the person's true intertemporal preferences U^t are time-consistent.[11] But when she incorrectly predicts how her tastes change over time, her perceived intertemporal preferences $\widetilde{U}^t$ can be time-inconsistent. Because this time inconsistency derives solely from misprediction of future utilities, it would make little sense to assume that the person is aware of it. We assume throughout the paper that the person is completely unaware of the time inconsistency — that at all times the person perceives her preferences to be time-consistent.[12]

Because perceived intertemporal preferences can change over time in a way that the person does not predict, projection bias can lead to *dynamic inconsistency* — she plans to behave a certain way in the future, but later, in the absence of new information, revises this plan. To highlight the nature of this inconsistency, we provide sufficient conditions for it not to occur:

Proposition 3.1. If a person has simple projection bias, then she will be dynamically consistent if for all $\mathbf{c}$ and $\mathbf{c}'$, $u(\mathbf{c},\mathbf{s}) - u(\mathbf{c}',\mathbf{s})$ is independent of $\mathbf{s}$.

Proposition 3.1 establishes that if the relative merits of any two consumption bundles — i.e., marginal utility — does not change over time, then projection bias cannot cause dynamic inconsistency. Proposition 3.1 therefore reveals that the source of dynamic inconsistency is mispredictions of future *marginal* utility, and not mispredictions of future utility *levels*. Consider, for instance, a student who starts smoking in college planning to quit upon graduation. She may underappreciate the extent to which smoking will decrease her quality of life, such as how it will lead to health problems. But as long as she correctly predicts how her craving for cigarettes will change over time,

[10] Research has, of course, documented a number of inadequacies of expected-utility theory (for an overview, see Starmer, 2000). To the extent that one feels the need to modify expected-utility theory for fully rational types, one could use the same modifications for people with projection bias.

[11] Another psychological phenomenon that has received increasing attention in research on intertemporal choice is hyperbolic discounting (see in particular, Laibson, 1994,1997, and O'Donoghue and Rabin, 1999*a*). Under hyperbolic discounting, true preferences are time-inconsistent.

[12] Given the logic of our model, it is inherent that a person is unaware of her *current* misprediction. But one could imagine a variant of the model where the person is aware of her *future* propensity to mispredict. She could, for instance, be aware of her general propensity to over-shop when hungry, while still committing the error on a case-by-case basis. The coexistence of day-to-day mispredictions with a "meta-awareness" of these mispredictions is similar to the discussion in O'Donoghue and Rabin (1999*b*) of how people can simultaneously be aware of their general tendency to procrastinate and yet still procrastinate on a case-by-case basis. A model of "sophisticated projection bias" could plausibly better describe behavior in some circumstances, but we choose our current formulation as a simple and realistic starting point.

she will be dynamically consistent — that is, she will carry out her plan to quit upon graduation. If, however, projection bias causes her to underestimate her future craving for cigarettes (increased marginal utility), she may continue to smoke after graduation contrary to her expectation and intention.

We believe that projection bias captures in a simple and tractable way a prevalent form of preference misprediction. Given any particular set of state-dependent preferences and particular economic environment, our model makes specific qualitative predictions about how actual behavior differs from fully rational behavior. Sections 4-6 highlight the potential importance of projection bias for economics by formally analyzing three specific environments.

4. Projection Bias and the Endowment Effect

As discussed in Section 2, a common experimental finding is the endowment effect — people tend to value objects more highly if they possess them than if they do not. The usual explanation of the endowment effect is loss aversion — people dislike losses relative to a reference point significantly more than they like gains. Loss aversion means that people tend to become attached to goods in their possession, and are reluctant to part with them, even if they would not have been willing to pay much to acquire them in the first place. Tversky and Kahneman (1991) show formally how loss aversion generates an endowment effect, and Strahilevitz and Loewenstein (1998) provide empirical support for the loss-aversion explanation of the endowment effect. In this section, we build a simple model of loss aversion and the endowment effect, and describe the implications of projection bias in this environment.

In the spirit of endowment-effect experiments, we frame our model in terms of preferences for an object such as a coffee mug. We consider two groups of people: (1) *sellers* are endowed with an object and have the option to sell it for money; and (2) *buyers* are not endowed with an object and have the option to buy one for money. The endowment effect is reflected in the finding that the reservation value for sellers is significantly larger than the reservation prices for buyers.

Suppose that in period t a person can either consume the object ($c_t = 1$) or not consume the object ($c_t = 0$). In addition, the person can either feel endowed ($s_t = 1$) or feel unendowed ($s_t = 0$). To formalize the typical experiment, we suppose that the decision whether to possess the object occurs in period 1, after which there is a second period during which the object, if possessed, can yield benefits.[13] If the person decides to possess the object, then $c_1 = c_2 = 1$; if the person decides not to possess the object, then $c_1 = c_2 = 0$. Both sellers and buyers must choose between these two consumption flows, but sellers feel endowed in period 1 — they have $s_1 = 1$ — and buyers feel unendowed — they have $s_1 = 0$.

We consider the following instantaneous utility function:

$$u(c_t, s_t) = \begin{cases} \mu + G \cdot (1 - s_t) & \text{if } c_t = 1 \\ \\ -L \cdot s_t & \text{if } c_t = 0. \end{cases}$$

In this formulation, the person receives intrinsic value μ from consuming the object. There is also a reference-dependent component to her utility function. If she consumes the object when she feels unendowed, she experiences a feeling of gain G; and if she does not consume the object when she feels endowed, then she experiences a feeling of loss L. To incorporate loss aversion, we assume $L > G$.

A person's feeling of endowment s_t comes from recent possession of the good. For simplicity, we assume that the person's reference level fully adjusts between periods 1 and 2, so that $s_2 = c_1$. Hence, if in period 1 the person chooses to possess the object, she will feel endowed in period 2 — she'll have $s_2 = 1$. Similarly, if in period 1 the person chooses not to possess the object, she will feel unendowed in period 2 — she'll have $s_2 = 0$.[14]

Finally, we assume no discounting, that any money received or paid enters as a separable and linear part of the intertemporal utility function, and that there is no reference dependence over money. Hence, a person with projection bias would choose to possess the object rather than receive (or pay) payment P only if $u(1, s_1) + \widetilde{u}(1, 1|s_1) \geq u(0, s_1) + \widetilde{u}(0, 0|s_1) + P$. It is straightforward to derive that the the reservation price for a buyer is $P^B(\alpha) \equiv 2\mu + (1 + \alpha)G$, and the reservation price for a seller is $P^S(\alpha) \equiv 2\mu + (1 + \alpha)L$.

The optimal (i.e., given $\alpha = 0$) reservation price for a buyer is $P^B(0) = 2\mu + G$, because if she buys the object, she experiences the intrinsic utility μ for two periods, experiences the feeling of gain G only in the first period, after which she adapts. Similarly, the optimal reservation price for

[13] Our qualitative conclusions crucially depend on there being at least one additional period in which the object yields benefits, since it is the future benefits (or forgone benefits) which the person mispredicts. But whether there is one additional period or many additional periods is not qualitatively important.

[14] A more general formulation is $s_2 = (1-\gamma)s_1 + \gamma c_1$ for some $\gamma \in (0, 1]$, where γ captures the speed of adaptation. This formulation of changing reference points is used in Ryder and Heal (1973), Bowman, Minehart, and Rabin (1999), and Strahilevitz and Loewenstein (1998). Our example assumes $\gamma = 1$, but the qualitative conclusions hold for any $\gamma \in (0, 1]$.

a seller is $P^S(0) = 2\mu + L$, because if she sells the object, she forgoes the intrinsic utility μ for two periods, but experiences the feeling of loss L only in the first period, after which she adapts. Projection bias leads to higher reservation prices for both buyers and sellers — P^B and P^S are both increasing in α. Intuitively, buyers overestimate the pleasure they'll feel from obtaining an object because they believe that they will continue to feel this pleasure further into the future than they actually will; and sellers overestimate the pain they'll feel from parting with an object because they believe that they will continue to feel this pain further into the future than they actually will. These conclusions are simple examples of a more general implication: Projection bias leads people to over-value reference-dependent goods.

A person exhibits an endowment effect if her selling price is larger than her buying price. Because $P^S(\alpha) > P^B(\alpha)$ for all α, the person exhibits an endowment effect whether or not she suffers from projection bias, which reflects that the endowment effect here is a manifestation of real preferences. If a person feels short-term gains and losses when she obtains and parts with objects, and if losses loom larger than gains, then even rational behavior leads to an endowment effect. By observing that $P^S(\alpha) - P^B(\alpha) = (1 + \alpha)(L - G)$ is increasing in α, however, we see that projection bias *magnifies* the endowment effect. Since projection bias leads a person to project her asymmetric responses to losses vs. gains into the future, it implies that the endowment effect is an exaggerated response to her real preferences.[15]

Finally, we note that, with the addition of an additional period 0 in which the person predicts her own future buying or selling price, it is easy to show that projection bias also predicts underestimation of buying prices by sellers and underprediction of selling prices by buyers, as found by Loewenstein and Adler (1995).

Our analysis above examines how projection bias changes an individual's reservation price; we next consider how projection bias might influence market outcomes. Consider a simple resale or second-hand market in which both buyers and sellers are price takers. We assume for simplicity that all people have the same degree of projection bias — they have the same α. We do assume, however, that people differ in terms of their intrinsic valuations for the object. Specifically, we assume there is a large population of buyers (a continuum with mass one), whose intrinsic valuations are distributed according to $F^B(\mu)$; and there is a large population of sellers (a continuum with

[15] Indeed, this implication that the endowment effect is in part an error is consistent with arguments by Kahneman (1991, p. 143) and Tversky and Kahneman (1991) that the endowment effect is a "bias" because people's actual pain when losing an object is not commensurate with their unwillingness to part with that object. Evidence from Strahilevitz and Loewenstein (1998) supports this interpretation, by showing that the speed of adaptation when people obtain and part with objects seems inconsistent with the magnitudes of the endowment effect usually observed.

mass one), whose intrinsic valuations are distributed according to $F^S(\mu)$. We assume $F^B(\mu)$ and $F^S(\mu)$ are both strictly increasing over support $[0, \mu^{\max}]$, where $\mu^{\max}$ may be infinite.[16]

Rearranging our conditions above, as a function of P, buyers demand the object when $\mu \geq \frac{P}{2} - \frac{(1+\alpha)G}{2}$, and sellers supply the object when $\mu \leq \frac{P}{2} - \frac{(1+\alpha)L}{2}$. Hence, the demand for objects is $D(P;\alpha) = 1 - F^B\left(\frac{P}{2} - \frac{(1+\alpha)G}{2}\right)$, and the supply for objects is $S(P;\alpha) = F^S\left(\frac{P}{2} - \frac{(1+\alpha)L}{2}\right)$. The equilibrium price $P^*(\alpha)$ satisfies $D(P^*(\alpha);\alpha) = S(P^*(\alpha);\alpha) \equiv Q^*(\alpha)$, where $Q^*(\alpha)$ is the equilibrium volume of trade.

Proposition 4.1. *P^* is strictly increasing in α, and Q^* is strictly decreasing in α.*

Proposition 4.1 establishes that, in a simple resale market, projection bias leads to increased prices and decreased volume of trade. Because projection bias makes both buyers and sellers more desirous of the object, it leads to both increased demand and decreased supply. These two effects unambiguously imply higher equilibrium prices — indeed, this result follows from either $L > 0$ or $G > 0$. The volume of trade falls because of loss aversion ($L > G$), which implies that the supply effect is larger than the demand effect. To illustrate the result in Proposition 4.1, we provide a closed-form solution for the case of a uniform distribution of buyers and sellers. If the distribution of buyers and the distribution of sellers are both uniform on $[0, \mu^{\max}]$, it is straightforward to derive that $P^*(\alpha) = \mu^{\max} + \frac{(1+\alpha)}{2}(L+G)$, and $Q^*(\alpha) = \frac{1}{2} - \frac{1+\alpha}{4\mu^{\max}}(L-G)$. Notice that P^* is increasing in α if either $L > 0$ or $G > 0$, whereas Q^* is decreasing in α only if $L > G$.

Our conclusions above apply to markets in which both buyers and sellers suffer similarly from projection bias. Because it is unlikely that firms or profit-oriented traders are likely to suffer much from projection bias, this assumption is most realistic in resale markets for houses, cars, wines, baseball cards, and so forth, where both sides of the market consist of individual consumers who value ownership of the objects. In markets where only one side of the market is likely to suffer from projection bias, the conclusions would change. If buyers have projection bias and sellers do not — e.g., consumers buying from competitive firms — projection bias affects only demand, and therefore leads to increased prices and increased quantities. If, in contrast, sellers have projection bias and buyers do not — e.g., consumers selling to professional buyers — projection bias affects only supply, and therefore leads to increased prices and decreased quantities.

Our analysis of market behavior above assumes price-taking behavior. In strategic interactions, an additional wrinkle arises: A person must predict what other people's reservation values are. In

[16] We also assume $\mu^{\max} > L - G$, which guarantees trade.

an extension of the study described in Section 2, Van Boven, Dunning, and Loewenstein (2000) demonstrate that such predictions are biased by an interpersonal projection bias, and as a result market efficiency suffers. Specifically, subjects were randomly assigned the role of "owner" or "buyer's agent". Owners were endowed with a coffee mug and stated a private minimum selling price. Buyers' agents were given \$10 to buy a mug for a principal (the experimenter). Buyers' agents were randomly paired with an owner and stated a buying price. If the price was higher than the owner's minimum selling price, then the owner received the amount offered by the buyer, the experimenter received the mug, and the buyer's agent kept the difference between \$10 and the buy price. If the price was lower, then the owner kept the mug and the buyer's agent gave back the full \$10. Consistent with projection bias, buyers' agents underestimated owners' minimum selling prices and bid too low, and as a result profits and volume were significantly below the profit-maximizing value. Van Boven, Loewenstein, and Dunning (forthcoming) show that the bias is not measurably reduced as a result of market experience, but is substantially reduced by simply giving the buyer's agent a mug to keep prior to the beginning of the study. Thus endowed, buyers' agents seemed better able to imagine how attached owners were to their mugs.[17]

5. Impulse Purchases of Durable Goods

People experience day-to-day changes in their tastes for many goods. Projection bias implies that they underappreciate such changes. When making short-term consumption decisions — whether to have eggplant vs. tofu at lunch right now — such mispredictions are irrelevant. But for long-term consumption decisions, such mispredictions can be important. In this section, we present a stylized model that identifies some possible implications of projection bias for one particular long-term consumption decision: whether to buy a durable good.

Consider a person who is deciding whether to buy a durable good, such as a tent, golf driver, or Johnny Depp video. The satisfaction that the person derives from the good might change over time for (at least) two reasons. First, her valuation is likely to systematically decline over time as the "novelty" of the item wears off. Second, there are likely random, day-to-day fluctuations in her

[17] Genesove and Mayer (2001) find evidence of financial loss aversion in housing markets — of people experiencing "pain" when they realize a nominal loss on their home. In particular, they find that sellers subject to nominal losses set higher asking prices and exhibit a lower hazard rate of sale. Our analysis in this section suggests that the magnitudes of these effects may be larger than justified by any true feelings of pain that people experience if and when they do sell their residence at a loss.

valuation. To formally model such effects, we assume the person's valuation of the good in period τ is

$$\phi_\tau \equiv \begin{cases} \mu_\tau & \text{if she does not yet own the good} \\ \gamma^{k-1}\mu_\tau & \text{if she purchased the good } k \leq D \text{ periods ago} \\ 0 & \text{if she purchased the good } k > D \text{ periods ago.} \end{cases}$$

In this formulation, μ_τ is a random variable that captures day-to-day fluctuations; we assume that μ_τ is independent across periods, that μ_τ has support $[\mu_L, \mu_H]$ and mean $\bar{\mu}$, and that the person learns the realization of μ_τ at the start of period τ. The term γ^{k-1} captures the systematic decline in the person's valuation, where $\gamma \in [0, 1]$ is a constant. We further assume that the durable good lasts for exactly D days. Finally, we assume that the person cannot consume the good on the day she purchases it.[18]

We first consider a situation in which the consumer has just one opportunity, on Day 1, to purchase the item. If she does not purchase it on Day 1, she cannot purchase it at all. We normalize the person's intertemporal utility to be zero when she does not buy the product. If she buys the product at price P, she will enjoy the benefits of ownership, but must forego the consumption of other goods that she could have financed with wealth P.[19] We assume that the person's utility from the durable good is additively separable from her utility for other goods, and that the price P represents the total utility value of the other goods forgone by purchasing the durable good. To fit the valuations described above within our framework, we let the person's consumption in period τ be $c_\tau = 1$ whenever she consumes the durable good in period τ, and $c_\tau = 0$ if not. The person's state in period τ is her current valuation, or $s_\tau = \phi_\tau$. Then her period-τ instantaneous utility from the durable good is $u(c_\tau, s_\tau) = c_\tau s_\tau$. Finally, we assume for simplicity that there is no discounting, or $\delta = 1$; none of our conclusions depend on this assumption.[20]

If the person buys the durable good in period 1, then, given the information available, her true expected intertemporal utility is

[18] Although this formulation puts restrictions on how the distribution of ϕ_τ varies with k, our results depend very little on the distribution of ϕ_τ. In particular, our results depend on only the mean of ϕ_τ and the support of ϕ_τ when $k = 1$. Also, while it is probably not realistic to assume that the person cannot consume the good on the day she purchases it, none of our qualitative conclusions depend on this assumption, and it vastly simplifies our analysis.

[19] We take the price P to be exogenous. In Loewenstein, O'Donoghue, and Rabin (2000), we formulate an alternative (more complicated) model with endogenous pricing by a monopolist, and permit the possibility of endogenously determined "sales hype" influencing the person's valuation.

[20] When it matters, we actually assume $\delta < 1$ and take the limit as $\delta \to 1$. In particular, if $\delta = 1$ implies the person is indifferent between buying now vs. buying in the future, then we say she prefers to buy now (because the preference is strict for δ near 1).

$$E_1\left[U^1\right] = E_1\left[\sum_{k=1}^{D}\gamma^{k-1}\mu_{1+k} - P\right] = \sum_{k=1}^{D}\gamma^{k-1}\bar{\mu} - P = \frac{1-\gamma^D}{1-\gamma}\bar{\mu} - P.$$

A person exhibiting simple projection bias perceives her expected intertemporal utility to be

$$\begin{aligned}
E_1\left[\tilde{U}^1\right] &= E_1\left[\sum_{k=1}^{D}\left[(1-\alpha)\gamma^{k-1}\mu_{1+k} + \alpha\mu_1\right] - P\right] \\
&= \sum_{k=1}^{D}\left[(1-\alpha)\gamma^{k-1}\bar{\mu} + \alpha\mu_1\right] - P = (1-\alpha)\frac{1-\gamma^D}{1-\gamma}\bar{\mu} + \alpha D\mu_1 - P.
\end{aligned}$$

Hence, the difference between the person's perceived intertemporal utility and her true intertemporal utility is

$$E_1\left[\tilde{U}^1\right] - E_1\left[U^1\right] = \alpha\left[D\bar{\mu} - \frac{1-\gamma^D}{1-\gamma}\bar{\mu}\right] + \alpha\left[D\mu_1 - D\bar{\mu}\right].$$

This equation reveals that projection bias creates two distortions in purchase decisions: The person underappreciates how her enjoyment of the good will diminish over time, which is reflected in the first term, and underappreciates day-to-day fluctuations, which is reflected in the second term.[21]

An underappreciation of "hedonic decay" unambiguously creates a tendency to over-buy durable goods, as can be seen by observing that $\frac{1-\gamma^D}{1-\gamma} < D$ for any $\gamma < 1$. Hence, the first distortion is always positive. Intuitively, if the person's true (expected) valuation of a durable good declines over time, but the person underappreciates the magnitude of this decline, she will over-value that durable good.

An underappreciation of day-to-day fluctuations can lead variously to under-buying or over-buying. Because the person extrapolates her valuation at purchase time into the future, she is too sensitive to her valuation at purchase time. Hence, if her day-1 valuation is larger than average, and she projects this above-average valuation onto the future, she is prone to over-value the durable good. If, in contrast, her day-1 valuation is smaller than average, and she projects this below-average valuation onto the future, she is prone to under-value the durable good. These conclusions are reflected in the second distortion being positive when $\mu_1 - \bar{\mu} > 0$ and negative when $\mu_1 - \bar{\mu} < 0$.[22]

[21] Our formulation assumes the same projection bias applies on both dimensions; we do not believe any non-obvious conclusions would be drawn from considering separate biases in predicting day-to-day fluctuations vs. decay in tastes.

[22] Because we assume that the person cannot consume the good on the day she purchases it, a rational type is entirely insensitive to her day-1 valuation. More generally, a rational type will be sensitive to her day-1 valuation, but a projector is still over-sensitive to her day-1 valuation — indeed, the conclusion generalizes that an underappreciation of day-to-day fluctuations leads a

Hence, when there are day-to-day fluctuations in how much a person enjoys a durable good, and the individual has only one exogenously determined opportunity to purchase the good, projection bias may make a person either more likely or less likely to buy a durable good. Things change dramatically, however, in the more realistic case where the person has multiple opportunities to buy the durable good. To make this point in a particularly stark way, we suppose that the consumer is going to purchase the good only once, and can buy the good in any period $t \in \{1, 2, ...\}$. In this situation, a rational person will either buy the durable good immediately in period 1 or never buy the durable good, and she buys the durable good if and only if $\frac{1-\gamma^D}{1-\gamma}\bar{\mu} - P \geq 0$. Intuitively, given our assumption that the person cannot consume the good on the day she purchases it, the net value of the durable good is independent of the purchase date. Hence, the good is either worth purchasing, in which case a rational person buys it immediately, or it is not worth purchasing, in which case she never buys it.

A person with projection bias, like a rational person, always perceives that the net value of the durable good is independent of the purchase date. As a result, she will purchase the good in the first period in which she perceives the good to be worth purchasing, which holds in period t if $(1-\alpha)\frac{1-\gamma^D}{1-\gamma}\bar{\mu} + \alpha D\mu_t - P \geq 0$. Assuming that μ_t is distributed with full support on $[\mu_L, \mu_H]$, there will eventually be some period in which the person perceives the good to be worth purchasing, and therefore the the person will (eventually) buy the good, if and only if $(1-\alpha)\frac{1-\gamma^D}{1-\gamma}\bar{\mu} + \alpha D\mu_H - P > 0$. Because $(1-\alpha)\frac{1-\gamma^D}{1-\gamma}\bar{\mu} + \alpha D\mu_H > \frac{1-\gamma^D}{1-\gamma}\bar{\mu}$, a person with projection bias is unambiguously more prone to buy the durable good than is a rational person. Hence, people who should buy will buy, but people who shouldn't buy may buy nevertheless.

The intuition behind this conclusion is an inherent asymmetry in purchases of durable goods. A decision not to buy is reversible, so if the person doesn't buy today when she should, she can still buy in the future. But a decision to buy is irreversible, so if she buys today when she shouldn't, she cannot un-buy in the future. With multiple buying opportunities, a person is prone not to buy when she should only in the unlikely event that she has a particularly low valuation *on every* buying opportunity, whereas she is prone to buy when she shouldn't in the quite likely event that she has a particularly high valuation *on at least one* buying opportunity.[23] Our main conclusion, therefore, is that, because people eventually have high valuations that they project onto their future utility, people are prone to over-buy durable goods. Hence, projection bias represents a source of

person to over-value (under-value) the good when $\mu_1 - \bar{\mu} > 0$ ($\mu_1 - \bar{\mu} < 0$).

[23] While a high-valuation day occurs with probability one in our extreme model, in real-world settings, the probability of experiencing an abnormally high valuation would be lower, and would depend on factors such as how long a store keeps a particular item in stock. But the asymmetry between buying and not buying is still present.

"impulse purchases" wherein people buy durable goods when it is not in their own self-interest.[24]

Our analysis suggests that certain types of sales tactics might be understood as attempts by businesses to exploit projection bias. If consumers exaggerate the longevity of feelings created by hot states, sellers will have an incentive to get people hot when they are making buying decisions. Such attempts at arousal might take the form of sales hype, enticing displays, or mood-inducing music. Sellers will also have an incentive to pressure people to make purchase decisions when hot, and to facilitate rapid purchases by consumers who are in a hot state that is unlikely to last, such as one-click shopping on the internet. Finally, projection bias might motivate firms to turn non-durable goods into durable goods via "intertemporal bundling" — e.g., selling memberships in health clubs, golf clubs, vacation time shares, and book clubs. Such practices allow firms to extract more surplus from consumers who suffer from projection bias. Consider, for instance, a person who becomes enthusiastic about exercise and makes a visit to a health club. Our model suggests that, rather than making a profit solely on that one visit, the health club may exploit the consumer's tendency to project her current enthusiasm into the future by offering a more expensive "club membership" that entitles the person to additional free (or low-cost) visits in the future. Indeed, DellaVigna and Malmendier (2001) empirically document that people over-pay for health club memberships. Using a panel data set that tracks members of three New England health clubs, they find that members who choose a contract with a flat monthly fee pay a price per visit of \$17, and members who choose a contract with a flat yearly fee pay a price per visit of \$15, even though a \$10-per-visit contract is also available. DellaVigna and Malmendier attribute these findings to naive self-control problems — people plan to attend frequently, but then don't have the self-control to carry out these plans. Our model suggests an alternative (perhaps complementary) explanation — people plan to attend frequently because they project their current enthusiasm into the future, but then decide not to attend in the future when their enthusiasm has waned.[25,26]

[24] Many examples of impulse purchases have (correctly) been attributed to other behavioral phenomena, perhaps most notably self-control problems (in the sense of hyperbolic discounting, as in Laibson 1994, 1997, O'Donoghue and Rabin 1999*a*). We believe that for durable goods, projection bias is a more important problem than hyperbolic discounting. Hyperbolic discounting provides a compelling explanation for over-consumption on cumulative small-scale consumption decisions, such as purchases of potato chips, where the net effects of repeated decisions to consume too many potato chips can be vast over-consumption of potato chips. But the purchase of a durable good is by its very nature a long-term-consumption decision — it's as if the person is choosing up front her total potato-chip consumption.

[25] DellaVigna and Malmendier also report evidence of costly delay in cancellation for contracts with automatic renewal, which cannot be explained by projection bias.

[26] Another possible explanation for their result is that people dislike paying on the margin for consumption (Prelec and Loewenstein, 1998). Prior research has observed a "flat-rate bias" in

In addition to helping to explain certain types of sales tactics, our model may also shed light on laws designed to counteract them. Cooling-off laws enacted at both the state and federal level allow consumers to rescind certain types of purchases within a few days of the transaction.[27] In the context of our model, such laws can be viewed as effective devices to combat the effects of projection bias. A cooling-off period that forces a consumer to reflect on her decision for several days can decrease the likelihood that she ends up owning a product that she shouldn't. Cooling-off laws may also have the additional benefit of reducing salespersons' incentives to hype if consumers can return products once they cool down and if such returns are costly for the seller.

6. Projection Bias and Habit Formation

Our final application of projection bias applies more directly to established economic models than our previous applications. For half a century, and more intensely recently, economists have explored life-cycle consumption models with habit formation. Habit formation — wherein increases in current consumption increase future marginal utility — was first proposed by Duesenberry (1949), and was first formalized by Pollak (1970) and Ryder and Heal (1973). In recent years, habit-formation models have been used in specific applications — see Becker and Murphy (1988), Constantinides (1990), Abel (1990), Campbell and Cochrane (1999), Jermann (1998), Boldrin, Christiano, and Fisher (2001), Carroll, Overland, and Weil (2000), and Fuhrer (2000). But these recent researchers have examined habit formation within the rational-choice framework. In this section, we describe the implications of habit formation in the presence of projection bias.[28]

We develop a simple "eat-the-cake" model wherein a person has income Y to allocate over consumption in periods $1, ..., T$, which we denote by $c_1, ..., c_T$. For simplicity, we assume that there is no discounting, and that the person can borrow and save at 0% interest; neither of these

domains such as internet and telephone billing plans. Kridel, Lehman, and Weisman (1993) find that about 65% of telephone customers who self-select flat rate service would have saved money by choosing a per-call billing option, but only 10% of those who selected variable rate service would have saved by choosing flat rate service. Neither projection bias nor hyperbolic time discounting seem like plausible explanations for this effect.

[27] For a detailed discussion of such laws, see Camerer, Issacharoff, Loewenstein, O'Donoghue, and Rabin (2002).

[28] The early literature on habit formation distinguishes between two polar cases: "rational habits" wherein consumers fully account for how current consumption affects future well-being, and "myopic habits" wherein consumers do not account at all for how current consumption affects future well-being. Of the papers cited in the text, all assume rational habits except for Pollak (1970), who (implicitly) assumes myopic habits. Muellbauer (1988) provides an excellent overview of the two extremes, and concludes that the empirical evidence seems to favor myopic habits. Our model is equivalent to rational habits when $\alpha = 0$ and to myopic habits when $\alpha = 1$.

assumptions is important for our qualitative conclusions. The person's true instantaneous utility in period t is $u(c_t, s_t)$, where the state s_t can usefully be thought of as her "habit stock". The person's initial habit stock, s_1, is exogenous, and her habit stock evolves according to $s_t = (1-\gamma)s_{t-1} + \gamma c_{t-1}$ for some $\gamma \in (0, 1]$. Hence, the more the person has consumed in the past, the higher is her current habit stock; and the more the person consumes now, the higher will be her future habit stock. The parameter γ represents how quickly the person develops (and eliminates) her habit.

We assume that instantaneous utility takes a particularly simple functional form:

$$u(c_t, s_t) = v(c_t - s_t) \qquad \text{where } v' > 0 \text{ and } v'' < 0.$$

This formulation is common in the literature, but is potentially quite restrictive.[29] Two features of this formulation are worth noting, both of which play a role in our results below. First, the marginal utility from consumption is increasing in the habit stock ($\partial[\partial u/\partial c]/\partial s > 0$), which implies habit formation — increases in current consumption increase the marginal utility from future consumption. Second, the level of utility is declining in the habit stock ($\partial u/\partial s < 0$), which implies that increases in current consumption reduce utility from future consumption. Although this "negative internality" is not an inherent part of habit formation, it is present in most formal analyses of — and in most real-world instances of — habit formation.[30]

In this environment, simple projection bias implies $\widetilde{u}(c, s|s') = (1-\alpha)v(c-s) + \alpha v(c-s')$. In period 1 the person faces the following choice problem, where s_1 is exogenous:

$$\begin{array}{lcl} \max_{(c_1,...,c_T)} \widetilde{U}^1(c_1, ..., c_T|s_1) & = & \sum_{\tau=1}^{T} \widetilde{u}(c_\tau, s_\tau|s_1) \\ \text{such that} & & s_t = (1-\gamma)s_{t-1} + \gamma c_{t-1} \text{ for } t \in \{2, ..., T\} \\ & & \text{and} \quad \sum_{\tau=1}^{T} c_\tau \leq Y. \end{array}$$

For ease of presentation below, let $(c_1^*, ..., c_T^*)$ denote optimal behavior, which solves this maximization when $\alpha = 0$, and let $(c_1^A, ..., c_T^A)$ denote planned behavior from the period-1 perspective for a person with $\alpha > 0$.[31]

Irrespective of α, there are three basic forces that influence behavior in this environment. First,

[29] This formulation is equivalent to that used by Pollak (1970), Constantanides (1990), Jermann (1998), Campbell and Cochrane (1999), and Boldrin, Christiano, and Fisher (2001); indeed, all these papers except Pollak further assume v takes a CRRA specification. Another formulation, proposed by Abel (1990) and used by Fuhrer (2000) and Carroll, Overland, and Weil (2000), is $u(c_t, s_t) = (c_t/s_t^\gamma)^{1-\sigma}/(1-\sigma)$. Yet a third formulation, suggested by Kahneman and Tversky's (1979) prospect theory, is to assume $v''(x) < 0$ for $x > 0$ but $v''(x) > 0$ for $x < 0$; Bowman, Minehart, and Rabin (1999) use a variant of this approach.

[30] The label "negative internality" comes from Herrnstein, Loewenstein, Prelec, and Vaughan (1993).

[31] We assume throughout that there are interior solutions for both optimal and actual behavior.

diminishing marginal utility yields the usual desire to smooth consumption. Indeed, when $\gamma = 0$, so that there is no habit formation, the person chooses a constant consumption profile of $c_t = Y/T$ for all t. Second, the negative internality creates a tendency to delay consumption. This can be seen most starkly when $v'' = 0$, meaning both constant marginal utility and no habit formation. In this case, a person would delay all consumption until period T.[32] The third force is more subtle, and reflects the effects of the person's habit stock changing over time due to her consumption decisions. If her habit stock is increasing over time, then her marginal utility is increasing over time, which creates a tendency to delay consumption; and if her habit stock is declining over time, then her marginal utility is declining over time, which creates a tendency to accelerate consumption.[33]

Lemma 6.1 describes how these forces interact for fully rational people:

Lemma 6.1. (1) If $c^*_\tau \geq s^*_\tau$ for some $\tau < T$, then $c^*_\tau < c^*_{\tau+1} < ... < c^*_T$. (2) If $s_1 = 0$, then $c^*_1 < c^*_2 < ... < c^*_T$.

Part 1 establishes that if there is ever a period in which a fully rational person consumes more than her habit stock, then she will follow an increasing consumption profile from that period onward. This result follows from two basic intuitions. First, from our discussion above, the only force that can lead a person to accelerate consumption is habit formation combined with a declining habit stock. In other words, a person will decrease consumption over time only if she is in the midst of a habit-breaking episode — an interval during which she is consuming below her habit stock so as to reduce her habit stock. Second, breaking a habit is both least painful and most beneficial when done early in life, before the habit has been further developed, and when the benefits will be spread over a large number of years. Hence, if the person is ever going to have a habit-breaking episode, it will be at the beginning of her life. Although this habit-breaking episode may last her whole life, if it ends, the person will thereafter follow an increasing consumption profile. Part 2 of Lemma 6.1 establishes that a sufficient condition for the optimal consumption profile to be globally increasing is $s_1 = 0$. This result follows directly from Part 1, because if the initial habit stock is zero then it is not possible to have an early-life habit-breaking episode.

We focus below on the implications of projection bias for situations in which optimal behavior

[32] People delay *all* consumption until period T because $v'' = 0$ eliminates the desire to consumption smooth. More generally, if $u(c, s) = v(c) + w(s)$ with $v' > 0$, $v'' < 0$, and $w' < 0$, which means there is diminishing marginal utility and no habit formation, people will choose $c_1 < ... < c_T$.

[33] One way to understand this third force is to consider an identical model except for having exogenous states, which would be a world in which the utility function changes over time in a fixed and predictable way. In that world, if $s_1 > s_2 > ... > s_T$ then people would choose $c_1 < c_2 < ... < c_T$, whereas if $s_1 < s_2 < ... < s_T$ then people would choose $c_1 > c_2 > ... > c_T$.

does not involve early-life habit-breaking episodes — that is, our results below only apply to parameter values such that $c_1^* \geq s_1$.[34] While $s_1 = 0$ is sufficient, a globally increasing consumption profile without habit-breaking episodes can arise more generally as long as $s_1 < Y/T$, which means the person's initial habit stock is sufficiently small that it is feasible to have $c_t > s_t$ for all t.[35]

Projection bias can give rise to two types of distortions in this environment, because the person underappreciates both the negative internality and the habit formation. The implication of an underappreciation of the negative internality is straightforward: Because the negative internality makes it optimal to delay consumption, projection bias makes the person prone to consume too much early in life and too little late in life relative to optimal behavior. The implication of an underappreciation of habit formation is in general somewhat complicated. But for the case in which optimal behavior does not involve a habit-breaking episode, and the person's habit stock is therefore increasing over time, habit formation also creates a tendency to delay consumption. Hence, an underappreciation of habit formation, like an underappreciation of the negative internality, makes the person prone to consume too much early in life and too little late in life relative to optimal behavior.[36] Proposition 6.1 reflects this intuition, establishing that whenever optimal behavior does not involve a habit-breaking episode, projection bias leads to over-consumption early in life and under-consumption late in life relative to what is optimal.

Proposition 6.1. If $c_1^* \geq s_1$, then $\sum_{t=1}^{\tau} c_t^A > \sum_{t=1}^{\tau} c_t^*$ for all $\tau < T$.

Our analysis above focuses on a person's period-1 plans. If the person does not suffer from projection bias, she will carry out those plans. But projection bias can lead the person's actual behavior to deviate from these plans as time passes and her tastes change in ways she did not predict. To study such effects, we examine how a person's plans change in period 2. In period 2, the person reoptimizes given her new perceived preferences — that is, she solves the following choice problem, where s_1 and c_1^A are exogenous:

[34] While $c_1^* \geq s_1$ is sufficient for our results below, we note that it does not guarantee that a person with projection bias will choose a globally increasing consumption profile. In most examples we have worked through, a projector chooses an increasing consumption profile when doing so is optimal; but we are able to construct examples where this is not the case, even when $s_1 = 0$.

[35] For $s_1 > Y/T$, the person *must* have a habit-breaking episode, and this episode might last her entire life — that is, she might have $c_1^* > c_2^* > ... > c_T^*$.

[36] When optimal behavior does involve a habit-breaking episode, during that episode habit formation is creating a tendency to accelerate consumption, and so an underappreciation of habit formation can create a tendency to consume too little early in life. Because this counteracts the effects of an underappreciation of the negative internality, the net effects of projection bias are ambiguous.

$$\begin{array}{lll} \max_{(c_2,...,c_T)} \widetilde{U}^1(c_2,...,c_T|s_2) & = & \sum_{\tau=2}^{T} \widetilde{u}(c_\tau, s_\tau|s_2) \\ \text{such that} & & s_2 = (1-\gamma)s_1 + \gamma c_1^A \\ & & s_t = (1-\gamma)s_{t-1} + \gamma c_{t-1} \text{ for } t \in \{3,...,T\} \\ & & \text{and} \quad \sum_{\tau=2}^{T} c_\tau \leq Y - c_1^A. \end{array}$$

Optimal behavior does not change over time, and hence the solution to this problem for $\alpha = 0$ is $(c_2^*, ..., c_T^*)$. For a person with projection bias, the solution for this problem, which we shall denote by $(c_2^{AA}, ..., c_T^{AA})$, may differ from her period-1 plans $(c_2^A, ..., c_T^A)$. Proposition 6.2 characterizes this revision of plans in the case where she is developing a habit and $T = 3$.

Proposition 6.2. Suppose $T = 3$ and $c_1^A > s_1$. Then $v''' > 0$ implies $c_2^{AA} > c_2^A$, $v''' < 0$ implies $c_2^{AA} < c_2^A$, and $v''' = 0$ implies $c_2^{AA} = c_2^A$.

As the person's habit stock changes over time, her (perceived) marginal utilities from consumption in each period also change. When the person is developing a habit, these marginal utilities all increase.[37] Hence, the relative magnitudes of these changes in marginal utility determine the revision of plans. If $v''' = 0$, then the increase in marginal utility is the same for all periods, which implies the person's marginal trade-offs have not changed, and hence she does not revise her consumption plan. If $v''' > 0$, then the increase in marginal utility is larger for period 2 than period 3, and as a result she revises her period-2 consumption upward. If $v''' < 0$, then the increase in marginal utility is smaller for period 2 and she revises her period-2 consumption downward.[38]

Any utility function that satisfies non-increasing absolute risk aversion, which includes the CARA and CRRA families, must have $v''' > 0$. Because this seems a plausible restriction on the instantaneous utility function, Proposition 6.2 suggests that projection bias leads people to repeatedly re-adjust their immediate consumption upwards relative to their most recent plans. Hence, if people experience habit formation in consumption, projection bias represents a possible source for actual saving being smaller than planned saving. There is considerable evidence that the

[37] Formally, from a period-t perspective, the (perceived) marginal utility from period-2 consumption is $(1-\alpha)v'(c_2 - s_2) + \alpha v'(c_2 - s_t) + (1-\alpha)\gamma v'(c_3 - s_3)$; and since $s_2 > s_1$ implies $v'(c_2 - s_2) > v'(c_2 - s_1)$, this marginal utility is larger from a period-2 perspective. Similarly, from a period-t perspective, the (perceived) marginal utility from period-3 consumption is $(1-\alpha)v'(c_3 - s_3) + \alpha v'(c_3 - s_t)$; and since $s_2 > s_1$ implies $v'(c_3 - s_2) > v'(c_3 - s_1)$, this marginal utility is also larger from a period-2 perspective.

[38] We conjecture, but have not proven, that this conclusion holds for $T > 3$. The result that $v''' = 0$ yields dynamic consistency is quite general. For the case $v''' > 0$, it is straightforward to show that marginal utility increases most for period 2 and least for period T, and so, perhaps subject to additional regularity conditions, after re-optimization we should expect period-2 consumption to increase and period-T consumption to decrease. Analogous conclusions hold for the $v''' < 0$ case.

actual saving of many households falls short of their plans (for an overview, see Laibson, Repetto, and Tobacman, 1998). The authors (and we ourselves) posit self-control problems as a primary source of this shortfall. But our analysis suggests that projection bias might also contribute to such mispredictions.

Our analysis above assumes that a person's lifetime income is exogenous. We conclude our analysis of projection bias and habit formation with a few thoughts on how projection bias might influence decisions about how hard to work to increase income. Rather than consider a full-blown model with endogenous labor-leisure decisions, we explore this topic more simply by considering what happens to the marginal utility of income in the model above. Let λ^A be the marginal utility of lifetime wealth as perceived from period 1, and let λ^{AA} be the marginal utility of lifetime wealth as perceived from period 2. Again limiting ourselves to the case when a person is developing a habit and the horizon is $T = 3$, Proposition 6.3 establishes that the marginal utility of income increases over time:

Proposition 6.3. Suppose $T = 3$ and $c_1^A > s_1$. Then $\lambda^{AA} > \lambda^A$.

Proposition 6.3 reflects a simple intuition: As time passes, and the person's real and perceived marginal utilities from consumption increase, income becomes more valuable. Extrapolating beyond our formal framework, this result suggests that projection bias would lead people to pursue higher income than planned as time passes. Projection bias would, for instance, create a force towards choosing a later and later planned retirement date as time passes, using the proceeds to increase consumption.

To illustrate this intuition, consider a person with complete projection bias $\alpha = 1$ who will live for 60 years and must choose both how many years to work and how to allocate her lifetime income. Each year that she works, she earns income \$1000. If in period t she consumes c_t, then her consumption utility is $u(c_t, s_t) = \ln(c_t - s_t + 25)$, where $s_1 = 425$ and the state in period $t > 1$ is $s_t = c_{t-1}$. Finally, every year after retirement she experiences additional "leisure utility" $\ell = 20$. In this environment, algebra shows that in period 1 she plans to work 27 years and consume \$450 per year, but then in period 2 she plans to work 29 years and consume \$484 per year for her remaining 59 years. In period 3 she will revise her plans again, to work 30 years and consume \$501 per year. This pattern continues, and in fact she will end up working all 60 years.

Similarly, if we were to introduce an endogenous per-period labor-leisure decision, projection bias over habit formation would create a tendency to repeatedly decide to increase labor and decrease leisure relative to earlier plans. We are wary of pushing this intuition too far without further theoretical and empirical analysis, however, because the logic of the argument assumes that there is no reference dependence in leisure. But we do note that this intuition parallels the arguments of many previous researchers, such as Scitovsky (1976) and Frank (1999), who have argued that people spend too much time and energy generating wealth and too little time on leisure activities, and that people enjoy increases in their standard of living less than they think they will.

7. Discussion and Conclusion

Our goal in this paper has been to improve the realism of the economic analysis of intertemporal choice by modelling a common form of misprediction of future preferences. The psychological evidence presented in Section 2 provides support for the existence of projection bias, and our analysis in Sections 4, 5, and 6 demonstrates the potential importance of projection bias in economics.

There are three reasons why projection bias should be incorporated into economic analysis. First, projection bias can explain certain phenomena, such as dynamic inconsistency, that are incompatible with the standard rational-choice model. Second, even when making similar comparative static predictions, projection bias may improve the *quantitative* behavioral predictions of economic models. In many contexts, the rational-choice model makes qualitatively correct predictions, but fails to make quantitatively plausible predictions. Just as Angeletos, Laibson, Repetto, Tobacman, and Weinberg (2001) have shown that hyperbolic discounting can provide a much better quantitative account than exponential discounting of savings behavior, in many environments projection bias may prove similarly useful. For instance, recent rational-choice models of addiction predict the directional effects of prices correctly (when prices go up, demand goes down), but seem to significantly exaggerate future-price elasticities. Projection bias may complement other explanations for addictive behavior to provide a more quantitatively accurate account of observed behavior. Finally, and perhaps most importantly, models incorporating projection bias can improve welfare analysis. By introducing a precise articulation of a systematic error in intertemporal-utility maximization, our model helps facilitate analysis of the ways in which people behave suboptimally in economic environments. Projection bias could, for instance, provide a principled way to study whether addicts, even price-sensitive ones, are making an optimal lifetime decision to become addicts.

How might one empirically identify projection bias in economic data? If only a single decision by each person can be observed, projection bias may be difficult to identify, because our model predicts that, at any point in time, the person behaves "rationally" in the sense that she has a well-defined set of predicted preferences and makes her decision to maximize those preferences. While our model says people may be wrong in their predicted utility, individuals would still obey the axioms of rational choice in one-shot decisions.[39] But projection bias can be inferred from one-shot decisions by large groups of people, using, for instance, field-data analogues of Read and van Leeuwen's (1998) experiments. If we observe two groups of people who must make decisions about consumption on the same future date, but are exogenously made to make the decisions in different circumstances, any difference in choices not plausibly attributed to different information would be attributable to biased predictions of future preferences.

Projection bias may also be identified if there is data with multiple observations for each decisionmaker, in which case researchers can examine whether behavior deviates from plans in a way predicted by projection bias. A direct approach is to ask people to report their plans, and then analyze how their behavior deviates from these plans. Some of the experimental evidence in Section 2, such as Loewenstein and Adler (1995), takes this approach. Another example of this approach is the survey evidence mentioned in Section 6 which shows that actual saving often falls short of people's plans. One might also infer plans from current decisions. An example of this approach is the health-club evidence discussed in Section 5: When people pay for a monthly health-club plan, one might infer that they plan to use the club sufficiently often to justify not paying on a daily basis. When other sources of dynamic inconsistency, such as hyperbolic discounting, can be ruled out in particular instances, then such dynamic inconsistency can be evidence for projection bias.

Sections 4, 5, and 6 outline the implications of projection bias in three specific economic environments, but we believe there are numerous other economic applications for which projection bias will prove important. An obvious application that we have alluded to repeatedly is addiction. Projection bias suggests that people might too often become addicted because they underappreciate the effects of current consumption on their own future preferences.[40] Our analysis — particularly in Section 5 — also suggests that people might overreact to transitory changes in the craving for addictive products. If, for instance, on a day when her craving is high the person overestimates her future desire for the drug, she may be discouraged from any efforts to quit. Analogously, if on a day when her craving is low the person underestimates her future desire for the drug, she may make a painful effort to quit — only to fail in this endeavor when her craving returns to average

[39] Kahneman (1994) distinguishes between "experienced utility", which reflects one's welfare, and "decision utility", which reflects the attractiveness of options as inferred from one's decisions; projection bias represents a reason why decision utility may deviate from experienced utility.

[40] Slovic (2001) finds quite good evidence that young people underappreciate the risk of becoming addicted if they indulge.

or high levels.[41]

A second application in a domain that is starting to receive attention among economists is to social-comparison theory, which studies the ways a person cares about her status relative to comparison groups. When people make decisions that cause their comparison groups to change — such as switching jobs or buying a house in a new neighborhood — projection bias predicts that people will underappreciate the effects of a change in comparison groups. As a result, people may be too prone to make reference-group-changing decisions that give them a sensation of status relative to their current reference group. If a person buys a small house in a wealthy neighborhood in part because it has a certain status value in her apartment building, she may not fully appreciate that her frame of reference may quickly become the larger houses and bigger cars that her new neighbors have.

Our review of evidence and our analysis in this paper leave some open questions. One is whether projection bias disappears with experience. That projection bias operates on states, such as hunger, with which people should have ample experience, suggests that projection bias does not disappear. Moreover, an explicit test of the effect of repeated experience failed to produce any appreciable learning. In the buyer's-agent studies discussed in Section 4, Van Boven, Loewenstein, and Dunning (forthcoming) offered buyers' agents five opportunities to bid for an object from sellers, with feedback after each bid about whether they had bid too high or too low. Bids, which were initially too low, did increase over the five rounds, and converged toward the profit-maximizing level. However, when a new object of approximately similar value was substituted for the original object, and another five rounds were conducted, initial bids began at the same low level as they had with the previous object, and the increase in price over the next five rounds was no more rapid as a result of subjects' earlier experience.

A related second open question is how aware are people of the bias. The existence of advice such as "count to ten before you respond" or "never shop on an empty stomach" suggests that people are aware of projection bias on a meta-level. In addition, we suspect that many rules people develop are designed to deal with moment-by-moment projection bias. For instance, in the context of our durable-good model, people might develop rules such as never buy a car on a first visit to a dealer. The need for such rules provides further evidence that people suffer from projection bias — but also implies that its damaging effects may be mitigated in many circumstances.

A third open question is the relationship between projection bias and diminishing marginal utility. Recent consumption of a product reduces the marginal utility from further consumption.

[41] See Loewenstein, O'Donoghue, and Rabin (2000) for an analysis of the role of projection bias in a simple two-period model of addiction.

Eating a second pint of ice cream yields less pleasure than the first, and watching a Johnny Depp movie for the 30^{th} time generates less pleasure than watching it for merely the 3^{rd} time. While consumer theory usually suppresses the temporal nature of diminishing marginal utility, it may be important once we recognize the existence of projection bias. Most people understand satiation: We all realize that eating the second pint of ice cream will be less satisfying than the first. But anecdotal evidence and an extrapolation of the hunger findings discussed earlier suggest projection bias leads people to underappreciate these effects. If people extrapolate marginal utilities in this way, then they will be prone to over-purchase activities they currently don't engage in. People may plan overly long vacations, believing the ninth day lying on the beach will be nearly as enjoyable as the first; and professionals who have little time for reading or traveling may falsely anticipate the blissfulness of spending their retirement years with non-stop reading and traveling. Firms may, of course, take advantage of such mispredictions, by selling large quantities in advance; restaurants may take advantage of projection bias by offering all-you-can-eat meals to hungry diners who underestimate how quickly they will become satiated.

As models that reflect the reality of both short-term fluctuations and long-term changes in preferences become more widespread in economics, economists must seriously address the question of whether people accurately predict how their preferences will change. Much as there has been a growing recognition among economists that behavioral and welfare economics will be improved by incorporating self-control problems into our models, we hope our analysis and examples illustrate the potential benefits for both behavioral and welfare economics of incorporating mispredictions of utilities in general, and projection bias in particular, into formal economic analysis.

Appendix A: More General Definition

Our analysis focuses entirely on simple projection bias, formalized in Definition 1. While simple projection bias is sufficient for the environments we consider in this paper, it is too restrictive for use as a general definition. In this appendix, we briefly describe two formulations of projection bias that allow for greater generality and realism.

Simple projection bias is problematic when there are multiple states, because it requires that the magnitude of the bias be identical for different types of states. For example, it requires that a person who is currently not thirsty and currently unaddicted to cocaine be just as bad at predicting her preferences when she is thirsty as she is at predicting her preferences when addicted to cocaine. This example suggests that one way to generalize Definition 1 is to apply simple projection bias on a state-by-state basis:

Definition 2. Suppose $\mathbf{s} \in \mathbb{R}^L$, and suppose there exist functions $v_1, ..., v_L$ such that $u(\mathbf{c}, \mathbf{s}) = \sum_{j=1}^{L} v_j(\mathbf{c}, s_j)$. Predicted utility exhibits *state-specific simple projection bias* if there exists $(\alpha_1, ..., \alpha_L) \in [0,1]^L$ such that for all $\mathbf{c}$, $\mathbf{s}$, and $\mathbf{s}'$, $\widetilde{u}(\mathbf{c}, \mathbf{s}|\mathbf{s}') = \sum_{j=1}^{L} \left[(1-\alpha_j)\ v_j(\mathbf{c}, s_j) + \alpha_j\ v_j(\mathbf{c}, s'_j) \right]$.

State-specific simple projection bias permits the magnitude of the bias to differ across different types of states. But this definition is problematic if states do not enter the utility function in an additively separable fashion. It also shares with simple projection bias a second problematic feature: the magnitude of the bias cannot depend on the current state. E.g., simple projection bias does not permit that a satiated person can predict well her preferences when hungry whereas a hungry person cannot predict well her preferences when satiated. The following further generalization permits this possibility while still imposing the idea that a person's predictions of future preferences are between her true future preferences and her current preferences.

Definition 3. Suppose $\mathbf{s} \in \mathbb{R}^L$, and let s_i denote its i^{th} element. We say $\mathbf{s}$ and $\mathbf{s}'$ differ only in element j if $s_j \neq s'_j$ and $s_i = s'_i$ for all $i \neq j$. Suppose $\mathbf{c} \in \mathbb{R}^K$, and let c_i denote its i^{th} element. For all $n \in \{1, 2, ...\}$, define $u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}) \equiv \frac{\partial^n u}{\partial c_{a_1} \partial c_{a_2} ... \partial c_{a_n}}(\mathbf{c}, \mathbf{s})$, where $a_i \in \{1, 2, ..., k\}$; these are all the n^{th}-order partial derivatives of the function $u(\mathbf{c}, \mathbf{s})$ with respect to the consumption variables. Define $\widetilde{u}^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}|\mathbf{s}')$ as the analogs of $u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s})$ for predicted utility, and define $u^0(\mathbf{c}, \mathbf{s}) = u(\mathbf{c}, \mathbf{s})$ and $\widetilde{u}^0(\mathbf{c}, \mathbf{s}|\mathbf{s}') \equiv \widetilde{u}(\mathbf{c}, \mathbf{s}|\mathbf{s}')$. We assume that $u(\mathbf{c}, \mathbf{s})$ and $\widetilde{u}(\mathbf{c}, \mathbf{s}|\mathbf{s}')$ are fully differentiable, so that all these items are well-defined. Finally, for any two real numbers x and y, let the set $G(x, y) \equiv [\min\{x, y\}, \max\{x, y\}]$ denote the interval between x and y.

Predicted utility exhibits *projection bias* if

(1) For all $\mathbf{c}$, $\mathbf{s}$ and $\mathbf{s}'$ such that $\mathbf{s}$ and $\mathbf{s}'$ differ only in element j, and for all $(n, a_1, a_2, ..., a_n)$, $\widetilde{u}^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}|\mathbf{s}') \in G\left(u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}), u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}')\right)$; and

(2) For all $\mathbf{c}$, $\mathbf{s}$, $\mathbf{s}'$, and $\mathbf{s}''$ such that $\mathbf{s}$, $\mathbf{s}'$, and $\mathbf{s}''$ differ in only element j, and for all $(n, a_1, a_2, ..., a_n)$, $u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}') \in G\left(u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}), u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}'')\right)$ implies $\widetilde{u}^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}|\mathbf{s}') \in G\left(u^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}), \widetilde{u}^n_{a_1 a_2 ... a_n}(\mathbf{c}, \mathbf{s}|\mathbf{s}'')\right)$.

Condition 1 says that, in addition to the predicted absolute level of utility being in between the true value and the current value, the various marginal utilities and cross-partials of all orders are also in between the true values and the current values. This implies that the person understands the qualitative nature of changes in her preferences, but underestimates the magnitudes of these changes. Condition 2 is a monotonicity property that says that the more the person's future preferences differ from her current preferences, the further her predictions are from her true future utility. Again, the condition says that not only is this true for the predicted absolute level of utility, but it is also true for the various marginal utilities and cross-partials. While none of the evidence we are familiar with directly implies this property, we feel it is a natural restriction

A generalization of Proposition 3.1 holds given this more general definition of projection bias:

Proposition 3.1′. A person will be dynamically consistent if for all $\mathbf{s}_t$, $\mathbf{s}_\tau$, $\mathbf{c}_\tau$, and $\mathbf{c}'_\tau$, $\widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}_t) - \widetilde{u}(\mathbf{c}'_\tau, \mathbf{s}_\tau|\mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_\tau) - u(\mathbf{c}'_\tau, \mathbf{s}_\tau)$, and for all $\mathbf{c}_\tau$, $\mathbf{s}_\tau$, $\mathbf{s}'_\tau$, $\mathbf{s}_t$, and $\mathbf{s}'_t$, $\widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}_t) - \widetilde{u}(\mathbf{c}_\tau, \mathbf{s}'_\tau|\mathbf{s}_t) = \widetilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}'_t) - \widetilde{u}(\mathbf{c}_\tau, \mathbf{s}'_\tau|\mathbf{s}'_t)$.

Proposition 3.1′ says that as long as projection bias does not cause a person either to misperceive the relative merits of any two consumption bundles, or to misperceive the relative impact on preferences of any two future states, then the person will be dynamically consistent.[42]

[42] The second condition necessarily holds for a simple projection bias.

Appendix B: Proofs

Proof of Proposition 3.1: For each superscript k, we let $(\mathbf{s}_1^k, ..., \mathbf{s}_T^k)$ denote the sequence of states induced by consumption path $\mathbf{C}^k \equiv (\mathbf{c}_1^k, ..., \mathbf{c}_T^k)$. Define $\tilde{V}^t(\mathbf{C}^k|\mathbf{s}_t^k) \equiv \sum_{\tau=t}^T \delta^{\tau-t}\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^k)$, which is the person's continuation utility when following consumption path $\mathbf{C}^k$ as perceived from period t. Suppose $\mathbf{C}^*$ is the optimal consumption path as perceived from period 1 — that is, $\mathbf{C}^*$ maximizes $\tilde{V}^1(\mathbf{C}|\mathbf{s}_1)$. The person will be dynamically consistent if for all $\mathbf{C}^k$ such that for some $\bar{\tau} > 1$, $\mathbf{c}_\tau^k = \mathbf{c}_\tau^*$ for all $\tau < \bar{\tau}$ (and therefore $\mathbf{s}_\tau^k = \mathbf{s}_\tau^*$ for all $\tau \leq \bar{\tau}$), $\tilde{V}^t(\mathbf{C}^*|\mathbf{s}_t^*) \geq \tilde{V}^t(\mathbf{C}^k|\mathbf{s}_t^*)$ for all $t \leq \bar{\tau}$.

Given $\mathbf{c}_\tau^k = \mathbf{c}_\tau^*$ for all $\tau < \bar{\tau}$ and $\mathbf{s}_\tau^k = \mathbf{s}_\tau^*$ for all $\tau \leq \bar{\tau}$, $\tilde{V}^t(\mathbf{C}^*|\mathbf{s}_t^*) - \tilde{V}^t(\mathbf{C}^k|\mathbf{s}_t^*) =$ $\sum_{\tau=\bar{\tau}}^T \delta^{\tau-t}\left[\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*)\right]$ for all $t \leq \bar{\tau}$. We prove $\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*)$ is independent of $\mathbf{s}_t^*$ for all $\tau \geq \bar{\tau}$, from which the result follows.

$\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*) = \left[\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*)\right] + \left[\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*)\right]$. $\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*) = (1-\alpha)\left[u(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*) - u(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*)\right] + \alpha\left[u(\mathbf{c}_\tau^*, \mathbf{s}_t^*) - u(\mathbf{c}_\tau^k, \mathbf{s}_t^*)\right] =$ $u(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*) - u(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*)$, which is independent of $\mathbf{s}_t^*$ (the last equality follows from the premise). $\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*) = (1-\alpha)\left[u(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*) - u(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k)\right]$, which is also independent of $\mathbf{s}_t^*$. The result follows.

Proof of Proposition 4.1: Given $D(P;\alpha) = 1 - F^B\left(\frac{P}{2} - \frac{(1+\alpha)G}{2}\right)$ and F^B strictly increasing, D is strictly decreasing in P and strictly increasing in α (on domain $[(1+\alpha)G, 2\mu^{\max}+(1+\alpha)G]$). Given $S(P;\alpha) = F^S\left(\frac{P}{2} - \frac{(1+\alpha)L}{2}\right)$ and F^S strictly increasing, S is strictly increasing in P and strictly decreasing in α (on domain $[(1+\alpha)L, 2\mu^{\max} + (1+\alpha)L]$). Note that $L > G$ and $\mu^{\max} > L - G$ guarantee that $Q^*(\alpha) \in (0,1)$ for all α.

To prove P^* is strictly increasing in α, posit otherwise. Then there exists α' and $\alpha'' > \alpha'$ such that $P^*(\alpha'') \leq P^*(\alpha')$. But then $D(P^*(\alpha'');\alpha'') > D(P^*(\alpha');\alpha')$ and $S(P^*(\alpha'');\alpha'') < S(P^*(\alpha');\alpha')$, and so it is not possible to have both $D(P^*(\alpha');\alpha') = S(P^*(\alpha');\alpha')$ and $D(P^*(\alpha'');\alpha'') = S(P^*(\alpha'');\alpha'')$. The result follows.

To prove Q^* is strictly decreasing in α, posit otherwise. Then there exists α' and $\alpha'' > \alpha'$ such that $Q^*(\alpha'') \geq Q^*(\alpha')$. Because $Q^*(\alpha) = D(P^*(\alpha);\alpha)$, $Q^*(\alpha'') \geq Q^*(\alpha')$ requires $\frac{P^*(\alpha'')-(1+\alpha'')G}{2} \leq \frac{P^*(\alpha')-(1+\alpha')G}{2}$ or $P^*(\alpha'') - P^*(\alpha') \leq (\alpha''-\alpha')G$. Similarly, because $Q^*(\alpha) = S(P^*(\alpha);\alpha)$, $Q^*(\alpha'') \geq Q^*(\alpha')$ requires $\frac{P^*(\alpha'')-(1+\alpha'')L}{2} \geq \frac{P^*(\alpha')-(1+\alpha')L}{2}$ or $P^*(\alpha'') - P^*(\alpha') \geq (\alpha''-\alpha')L$. But $L > G$ implies it is not possible to satisfy both inequalities. The result follows.

Proof of Lemma 6.1: To ease our notation, we use $v_t^* \equiv v'(c_t^* - s_t^*)$ for all t. Also, for any function $g(i)$, we say $\sum_{i=a}^{b} g(i) = 0$ when $a > b$.

(1) Given $\alpha = 0$, the first-order conditions are $v_t^* - X_t^* = \lambda^*$ for all t, where λ^* is the multiplier on the income constraint, and $X_t^* \equiv \gamma \sum_{\tau=t+1}^{T} (1-\gamma)^{\tau-(t+1)} v_\tau^*$. Hence, for all t, $v_{t-1}^* - X_{t-1}^* = v_t^* - X_t^*$ or $v_{t-1}^* - v_t^* = X_{t-1}^* - X_t^*$. Because $X_{t-1}^* - X_t^* = \gamma(v_t^* - X_t^*)$, and because $v_t^* - X_t^* = v_T^*$, it follows that for all t, $v_{t-1}^* - v_t^* = \gamma v_T^* > 0$, which in turn implies $v_t^* = (1 + (T-t)\gamma) v_T^*$. Hence, $v_1^* > ... > v_T^*$, which given $v'' < 0$ implies $c_1^* - s_1^* < ... < c_T^* - s_T^*$.

Now suppose $c_\tau^* \geq s_\tau^*$, in which case $s_{\tau+1}^* \geq s_\tau^*$, and for all $\tau > t$, $c_t^* > s_t^*$ and therefore $s_{t+1}^* > s_t^*$. Because for any t, $s_{t+1}^* \geq s_t^*$ and $c_{t+1}^* - s_{t+1}^* > c_t^* - s_t^*$ together imply $c_{t+1}^* > c_t^*$, the result follows.

(2) Because $s_1 = 0$ guarantees $c_1^* \geq s_1$, the result follows directly from part 1.

Proof of Proposition 6.1: We use v_t^* as in the proof of Lemma 6.1, and note that $c_1^* \geq s_1$ implies $c_1^* < ... < c_T^*$ and also $c_t^* - s_t^* > 0$ for all $t > 1$. We also use $v_t^A \equiv v'(c_t^A - s_t^A)$ and $\hat{v}_t \equiv v'(c_t^A - s_1)$, and note that $\hat{v}_t > \hat{v}_s$ if and only if $c_t^A < c_s^A$. The first-order conditions are $v_t^A - X_t^A + \frac{\alpha}{1-\alpha}\hat{v}_t = \lambda^A/(1-\alpha)$, where λ^A is the multiplier on the income constraint, and $X_t^A \equiv \gamma \sum_{\tau=t+1}^{T} (1-\gamma)^{\tau-(t+1)} v_\tau^A$. Hence, for all t, $v_{t-1}^A - v_t^A = X_{t-1}^A - X_t^A + \frac{\alpha}{1-\alpha}[\hat{v}_t - \hat{v}_{t-1}]$. Because $X_{t-1}^A - X_t^A = \gamma(v_t^A - X_t^A)$, and because $v_t^A - X_t^A = v_T^A + \frac{\alpha}{1-\alpha}[\hat{v}_T - \hat{v}_t]$, it follows that for all t, $v_{t-1}^A - v_t^A = \gamma v_T^A + \frac{\alpha}{1-\alpha}[\gamma \hat{v}_T - (\hat{v}_{t-1} - (1-\gamma)\hat{v}_t)]$. By starting with the condition for $t = T$ and iterating backwards, we can derive that for all t,
$v_t^A = (1 + (T-t)\gamma) v_T^A + \frac{\alpha}{1-\alpha}\left[(1 + (T-t)\gamma)\hat{v}_T - (\hat{v}_t + \gamma \sum_{i=t+1}^{T} \hat{v}_i)\right]$. It is useful to rewrite this condition as $\frac{v_t^A}{(1+(T-t)\gamma)} + \frac{\alpha}{1-\alpha}\frac{R_t}{(1+(T-t)\gamma)} = v_T^A + \frac{\alpha}{1-\alpha}\hat{v}_T$ where $R_t = (\hat{v}_t + \gamma \sum_{i=t+1}^{T} \hat{v}_i)$. Also note that for all t and s,

$$\frac{v_t^A - v_t^*}{(1 + (T-t)\gamma)} + \frac{\alpha}{1-\alpha}\frac{R_t}{(1 + (T-t)\gamma)} = \frac{v_s^A - v_s^*}{(1 + (T-s)\gamma)} + \frac{\alpha}{1-\alpha}\frac{R_s}{(1 + (T-s)\gamma)}. \tag{1}$$

We next establish two claims.

Claim 1: There exists t and s such that $v_t^A < v_t^*$ and $v_s^A > v_s^*$.

Proof: Suppose otherwise. First consider the case in which $v_t^A = v_t^*$ for all t, which implies $c_t^A = c_t^*$ for all t. Applying equation (1), $v_t^A = v_t^*$ for all t implies $R_t/(1+(T-t)\gamma) = R_s/(1+(T-s)\gamma)$ for all t and s; but this requires $\hat{v}_t = \hat{v}_s$ and therefore $c_t^A = c_s^A$ for all t and s, which contradicts $c_1^* < ... < c_T^*$. Next consider the case in which $v_t^A \leq v_t^*$ and therefore $c_t^A - s_t^A \geq c_t^* - s_t^*$ for all t, where the inequalities are strict for some t. For any t, if $s_t^A \geq s_t^*$, then $c_t^A - s_t^A \geq c_t^* - s_t^*$ implies

$c_t^A \geq c_t^*$ and therefore $s_{t+1}^A \geq s_{t+1}^*$, where either $s_t^A > s_t^*$ or $c_t^A - s_t^A > c_t^* - s_t^*$ implies $c_t^A > c_t^*$ and $s_{t+1}^A > s_{t+1}^*$. In addition, $c_1^A - s_1^A \geq c_1^* - s_1^*$ implies $c_1^A \geq c_1^*$ and therefore $s_2^A \geq s_2^*$, where $c_1^A - s_1^A > c_1^* - s_1^*$ implies $c_1^A > c_1^*$ and $s_2^A > s_2^*$. It follows that $c_t^A \geq c_t^*$ for all t and $c_t^A > c_t^*$ for some t, which contradicts that $\sum_{t=1}^T c_t^A = \sum_{t=1}^T c_t^* = Y$. Finally, an analogous logic rules out the case in which $v_t^A \geq v_t^*$ for all t and $v_t^A > v_t^*$ for some t.

Claim 2: There exists $\bar{\tau} \in \{1, ..., T-1\}$ such that $v_t^A \leq v_t^*$ for $t \in \{1, ..., \bar{\tau}\}$ and $v_t^A > v_t^*$ for $t \in \{\bar{\tau}+1, ..., T\}$.

Proof: Suppose otherwise. Let $x \equiv \max\left\{t \mid v_t^A \leq v_t^*\right\}$, which exists given Claim 1, and let $z \equiv \max\left\{t < x \mid v_t^A > v_t^*\right\}$, which must exist if the Claim 2 is not true. Applying equation (1), $v_z^A > v_z^*$ and $v_{z+1}^A \leq v_{z+1}^*$ together imply $R_z/(1+(T-z)\gamma) < R_{z+1}/(1+(T-z-1)\gamma)$, which means $\left[\hat{v}_z + \gamma\hat{v}_{z+1} + \gamma\sum_{i=z+2}^T \hat{v}_i\right]/(1+(T-z)\gamma) < \left[\hat{v}_{z+1} + \gamma\sum_{i=z+2}^T \hat{v}_i\right]/(1+(T-z-1)\gamma)$ or

$$[1+(T-z-1)\gamma]\,\hat{v}_z - [1+(T-z-1)\gamma(1-\gamma)]\,\hat{v}_{z+1} < \gamma^2 \sum_{i=z+2}^T \hat{v}_i. \tag{2}$$

We prove that inequality (2) cannot hold, from which Claim 2 follows.

We first establish that $c_t^A > c_z^A$ and therefore $\hat{v}_t < \hat{v}_z$ for all $t \in \{z+1, ..., x\}$. Because $v_t^* > v_{t+1}^*$ for all t, it follows that $v_z^A > v_z^* > v_t^* \geq v_t^A$ for all $t \in \{z+1, ..., x\}$. Since $v_t^A \leq v_t^*$ implies $c_t^A - s_t^A \geq c_t^* - s_t^*$, $c_t^* - s_t^* > 0$ implies $c_t^A > s_t^A$ and therefore $s_{t+1}^A > s_t^A$, and so $s_t^A > s_{z+1}^A$ for all $t \in \{z+2, ..., x\}$. If $c_z^A < s_z^A$ then $s_{z+1}^A > c_z^A$ and therefore $c_t^A > s_t^A \geq s_{z+1}^A > c_z^A$. If instead $c_z^A \geq s_z^A$ then $s_{z+1}^A \geq s_z^A$, and since $v_z^A > v_t^A$ implies $c_z^A - s_z^A < c_t^A - s_t^A$, $s_t^A \geq s_{z+1}^A \geq s_z^A$ implies $c_t^A > c_z^A$.

If $x = T$, then $\hat{v}_t < \hat{v}_z$ for all $t \in \{z+2, ..., T\}$ and therefore $\gamma^2\sum_{i=z+2}^T \hat{v}_i < \gamma^2(T-z-1)\hat{v}_z$. But then $\hat{v}_z > \hat{v}_{z+1}$ implies $\gamma^2(T-z-1)\hat{v}_z < [1+(T-z-1)\gamma]\,\hat{v}_z - [1+(T-z-1)\gamma(1-\gamma)]\,\hat{v}_{z+1}$, which contradicts inequality (2).

Consider instead $x < T$. Given $(v_t^A - v_t^*) + \frac{\alpha}{1-\alpha}R_t = (1-(T-t)\gamma)\left[v_T^A - v_T^* + \frac{\alpha}{1-\alpha}\hat{v}_T\right]$, it follows that for all t and s, $\frac{1}{s-t}\left[(v_t^A - v_t^*) - (v_s^A - v_s^*) + \frac{\alpha}{1-\alpha}(R_t - R_s)\right] = -\gamma\left[v_T^A - v_T^* + \frac{\alpha}{1-\alpha}\hat{v}_T\right]$. Hence, $(v_z^A - v_z^*) - (v_{z+1}^A - v_{z+1}^*) + \frac{\alpha}{1-\alpha}(R_z - R_{z+1}) = \frac{1}{n}\left[(v_x^A - v_x^*) - (v_{x+n}^A - v_{x+n}^*) + \frac{\alpha}{1-\alpha}(R_x - R_{x+n})\right]$. Given $v_z^A > v_z^*$, $v_{z+1}^A \leq v_{z+1}^*$, $v_x^A \leq v_x^*$, and $v_{x+n}^A > v_{x+n}^*$, it follows that $(R_x - R_{x+n}) - n(R_z - R_{z+1}) > 0$. Because $R_x - R_{x+n} = \hat{v}_x + \gamma\sum_{i=1}^{n-1}\hat{v}_{x+i} - (1-\gamma)\hat{v}_{x+n}$ and $R_z - R_{z+1} = \hat{v}_z - (1-\gamma)\hat{v}_{z+1}$, this condition becomes $(R_x - R_{x+n}) - n(R_z - R_{z+1}) =$
$(1-\gamma)(\hat{v}_{z+1} - \hat{v}_{x+n}) + [\hat{v}_x + (n-1)\hat{v}_{z+1} - n\hat{v}_z] + \gamma\sum_{i=1}^{n-1}(\hat{v}_{x+i} - \hat{v}_{z+1}) > 0$. Since $\hat{v}_x < \hat{v}_z$ and $\hat{v}_{z+1} < \hat{v}_z$, applying this condition for $n = 1$ yields $\hat{v}_{z+1} > \hat{v}_{x+1}$, and then applying it for $n = 2$ yields $\hat{v}_{z+1} > \hat{v}_{x+2}$, and so forth. It follows that $\hat{v}_{x+n} < \hat{v}_{z+1} < \hat{v}_z$ for all $n \in \{1, ..., T-x\}$, and therefore $\hat{v}_t < \hat{v}_z$ for all $t \in \{z+1, ..., T\}$. But then $\gamma^2\sum_{i=z+2}^T \hat{v}_i < \gamma^2(T-z-1)\hat{v}_z <$

$[1+(T-z-1)\gamma]\,\hat{v}_z - [1+(T-z-1)\gamma(1-\gamma)]\,\hat{v}_{z+1}$, which contradicts inequality (2). Claim 2 follows.

Finally, we prove the main result. Posit otherwise, and define $w \equiv \min\left\{\tau \;\middle|\; \sum_{i=1}^{\tau} c_i^A \leq \sum_{i=1}^{\tau} c_i^*\right\}$. Claims 1 and 2 together imply $v_1^A < v_1^*$, and therefore $c_1^A > c_1^*$. Hence, $w > 1$ and $c_w^A < c_w^*$. Note that if $w \leq \bar{\tau}$ (where $\bar{\tau}$ defined as in Claim 2) then $v_1^A < v_1^*$ and $v_t^A \leq v_t^*$ for all $t \in \{2, ..., w-1\}$, which implies $s_w^A > s_w^*$ (using logic identical to that in proof of Claim 1). But then $c_w^A < c_w^*$ implies $v_w^A > v_w^*$, which contradicts that $w \leq \bar{\tau}$. It follows that $w > \bar{\tau}$ and therefore $v_w^A > v_w^*$.

Define $y \equiv \min\left\{\tau > w \;\middle|\; c_\tau^A \geq c_\tau^*\right\}$; such a y must exist. We can write the state s_t as

$$\begin{aligned} s_t &= \gamma c_{t-1} + (1-\gamma)\gamma c_{t-2} + (1-\gamma)^2\gamma c_{t-3} + ... + (1-\gamma)^{t-2}\gamma c_1 + (1-\gamma)^{t-1} s_1 \\ &= \gamma \sum_{i=1}^{t-1} c_i - \gamma^2 \sum_{j=1}^{t-2}\left[(1-\gamma)^{j-1} \sum_{i=1}^{t-1-j} c_i\right] + (1-\gamma)^{t-1} s_1. \end{aligned}$$

Then $\sum_{i=1}^{w} c_i^* \geq \sum_{i=1}^{w} c_i^A$ and $\sum_{i=1}^{\tau} c_i^* < \sum_{i=1}^{\tau} c_i^A$ for all $\tau < w$ together imply

$$s_{w+1}^* - s_{w+1}^A = \gamma\left(\sum_{i=1}^{w} c_i^* - \sum_{i=1}^{w} c_i^A\right) - \gamma^2 \sum_{j=1}^{w-1}\left[(1-\gamma)^{j-1}\left(\sum_{i=1}^{w-j} c_i^* - \sum_{i=1}^{w-j} c_i^A\right)\right] > 0.$$

Moreover, when $y > w+1$, $s_{w+1}^* > s_{w+1}^A$ combined with $c_t^* > c_t^A$ for all $t \in \{w+1, ..., y-1\}$ implies $s_y^* > s_y^A$. Since, by the definition of y, $c_y^A \geq c_y^*$, it follows that $c_y^A - s_y^A > c_y^* - s_y^*$ and therefore $v_y^A < v_y^*$. But given $v_w^A > v_w^*$, this contradicts Claim 2. The result follows.

Proof of Proposition 6.2: As a preliminary step, we prove $c_1^A < c_3^A$ and $c_2^A < c_3^A$. Posit otherwise, and suppose $z \in \arg\max_{t\in\{1,2\}} c_t^A$. Hence, $c_z^A \geq c_{z+1}^A$ and $c_z^A \geq c_T^A$, which implies $\hat{v}_z \leq \hat{v}_{z+1}$ and $\hat{v}_z \leq \hat{v}_T$. Recall that $v_z^A - v_{z-1}^A = \gamma v_T^A + \frac{\alpha}{1-\alpha}\left[\gamma\hat{v}_T - (\hat{v}_z - (1-\gamma)\hat{v}_{z+1})\right]$. Because $\gamma\hat{v}_T - (\hat{v}_z - (1-\gamma)\hat{v}_{z+1}) = (1-\gamma)(\hat{v}_{z+1} - \hat{v}_z) + \gamma(\hat{v}_T - \hat{v}_z) \geq 0$, $v_z^A - v_{z+1}^A > 0$. Given $v'' < 0$, this implies $c_z^A - s_z^A < c_{z+1}^A - s_{z+1}^A$, and given $c_z^A \geq c_{z+1}^A$ this holds only if $s_z^A > s_{z+1}^A$, which in turn holds only if $c_z^A < s_z^A$. But if $z = 1$ this contradicts $c_1^A > s_1$, and if $z = 2$ this contradicts $c_2^A > c_1^A > s_2^A$ ($c_1^A > s_1$ implies $c_1^A > s_2^A$).

In period 1, true utility is $U^1(c_1, c_2, c_3) = \sum_{\tau=1}^{3} v(c_\tau - s_\tau)$, and perceived utility is $\widetilde{U}^1(c_1, c_2, c_3|s_1) = \sum_{\tau=1}^{3}\left[(1-\alpha)v(c_\tau - s_\tau) + \alpha v(c_\tau - s_1)\right] =$ $(1-\alpha)U^1(c_1, c_2, c_3) + \alpha\sum_{\tau=1}^{3} v(c_\tau - s_1)$. Period-1 behavior (c_1^A, c_2^A, c_3^A) must satisfy $\frac{\partial\widetilde{U}^1(c_1^A,c_2^A,c_3^A|s_1)}{\partial c_1} = \frac{\partial\widetilde{U}^1(c_1^A,c_2^A,c_3^A|s_1)}{\partial c_2} = \frac{\partial\widetilde{U}^1(c_1^A,c_2^A,c_3^A|s_1)}{\partial c_3}$. Because $\frac{\partial\widetilde{U}^1(c_1,c_2,c_3|s_1)}{\partial c_t} = (1-\alpha)\frac{\partial U^1(c_1,c_2,c_3)}{\partial c_t} + \alpha v'(c_t - s_1)$ for $t \in \{1, 2, 3\}$, $\frac{\partial\widetilde{U}^1(c_1^A,c_2^A,c_3^A|s_1)}{\partial c_2} = \frac{\partial\widetilde{U}^1(c_1^A,c_2^A,c_3^A|s_1)}{\partial c_3}$ implies $(1-\alpha)\left[\frac{\partial U^1(c_1^A,c_2^A,c_3^A)}{\partial c_2} - \frac{\partial U^1(c_1^A,c_2^A,c_3^A)}{\partial c_3}\right] = \alpha\left[v'(c_3^A - s_1) - v'(c_2^A - s_1)\right]$.

After choosing $c_1^A > s_1$, in period 2, the state is $s_2^A = (1-\gamma)s_1 + \gamma c_1^A$, true utility is $U^2(c_2, c_3|s_2^A) =$

$\sum_{\tau=2}^{3} v(c_\tau - s_\tau)$, and perceived utility is $\widetilde{U}^2(c_2, c_3|s_2^A) = \sum_{\tau=2}^{3}\left[(1-\alpha)v(c_\tau - s_\tau) + \alpha v(c_\tau - s_2^A)\right] = (1-\alpha)U^2(c_2, c_3|s_2^A) + \alpha\sum_{\tau=2}^{3} v(c_\tau - s_2^A)$. Period-2 behavior (c_2^{AA}, c_3^{AA}) must satisfy $\frac{\partial\widetilde{U}^2(c_2^{AA}, c_3^{AA}|s_2^A)}{\partial c_2} = \frac{\partial\widetilde{U}^2(c_2^{AA}, c_3^{AA}|s_2^A)}{\partial c_3}$. Note that for $t \in \{2,3\}$, $\frac{\partial U^1(c_1^A, c_2, c_3)}{\partial c_t} = \frac{\partial U^2(c_2, c_3|s_2^A)}{\partial c_t}$ for all c_2 and c_3. Hence, because $\frac{\partial\widetilde{U}^2(c_2, c_3|s_2^A)}{\partial c_t} = (1-\alpha)\frac{\partial U^1(c_1^A, c_2, c_3)}{\partial c_t} + \alpha v'(c_t - s_2^A)$ for $t \in \{2,3\}$, $\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^A)}{\partial c_2} - \frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^A)}{\partial c_3} = \alpha\left[v'(c_3^A - s_1) - v'(c_2^A - s_1)\right] + \alpha\left[v'(c_2^A - s_2^A) - v'(c_3^A - s_2^A)\right]$.

$v''' > 0$, $s_2^A > s_1$ (which follows from $c_1^A > s_1$), and $c_2^A < c_3^A$ together imply $v'(c_2^A - s_2^A) - v'(c_2^A - s_1) > v'(c_3^A - s_2^A) - v'(c_3^A - s_1)$, which in turn implies $\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2)}{\partial c_2} > \frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2)}{\partial c_3}$. Given the concavity of $\widetilde{U}^2$, we must have $c_2^{AA} > c_2^A$ and $c_3^{AA} < c_3^A$.

An analogous argument holds for $v''' < 0$.

$v''' = 0$ implies $v'(c_2^A - s_2^A) - v'(c_2^A - s_1) = v'(c_3^A - s_2^A) - v'(c_3^A - s_1) = k(s_2 - s_1)$ for some constant k (i.e., $v''' = 0$ implies v' is linear and decreasing, so $-k$ is the slope of v'), and so $\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2)}{\partial c_2} = \frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2)}{\partial c_3}$. It follows that $(c_2^{AA}, c_3^{AA}) = (c_2^A, c_3^A)$. (The conclusion that $v''' = 0$ yields dynamic consistency would hold for any T and for any c_1^A.)

Proof of Proposition 6.3: Using the notation from the proof of Proposition 6.2, $\lambda^A = \frac{\partial\widetilde{U}^1(c_1^A, c_2^A, c_3^A|s_1)}{\partial c_1} = \frac{\partial\widetilde{U}^1(c_1^A, c_2^A, c_3^A|s_1)}{\partial c_2} = \frac{\partial\widetilde{U}^1(c_1^A, c_2^A, c_3^A|s_1)}{\partial c_3}$, and $\lambda^{AA} = \frac{\partial\widetilde{U}^2(c_2^{AA}, c_3^{AA}|s_2^A)}{\partial c_2} = \frac{\partial\widetilde{U}^2(c_2^{AA}, c_3^{AA}|s_2^A)}{\partial c_3}$. The concavity of $\widetilde{U}^2$ implies that $\lambda^{AA} \geq \min\left\{\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^A)}{\partial c_2}, \frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^A)}{\partial c_3}\right\}$. For $t \in \{2,3\}$ we have $\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^A)}{\partial c_t} = \frac{\partial\widetilde{U}^1(c_1^A, c_2^A, c_3^A|s_1)}{\partial c_t} + \alpha\left[v'(c_t^A - s_2^A) - v'(c_t^A - s_1)\right]$. Then $s_2^A > s_1$ (which follows from $c_1^A > s_1$) combined with $v'' < 0$ implies $v'(c_t^A - s_2^A) > v'(c_t^A - s_1)$. Hence, for $t \in \{2,3\}$, $\frac{\partial\widetilde{U}^2(c_2^A, c_3^A|s_2^2)}{\partial c_t} > \frac{\partial\widetilde{U}^1(c_1^A, c_2^A, c_3^A|s_1)}{\partial c_t} = \lambda^A$. The result follows.

Proof of Proposition 3.1$'$: As in the proof of Proposition 3.1, the result follows if $\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*)$ is independent of $\mathbf{s}_t^*$ for all $\tau \geq \bar{\tau}$.

$\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*) = \left[\tilde{u}(\mathbf{c}_\tau^*, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*)\right] + \left[\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^*|\mathbf{s}_t^*) - \tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k|\mathbf{s}_t^*)\right]$. The first condition in Proposition 3.1$'$ implies $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau|\mathbf{s}_t) - \tilde{u}(\mathbf{c}'_\tau, \mathbf{s}_\tau|\mathbf{s}_t)$ is independent of $\mathbf{s}_t$. The second condition of the Proposition 3.1$'$ implies $\tilde{u}(\mathbf{c}'_\tau, \mathbf{s}_\tau|\mathbf{s}_t) - \tilde{u}(\mathbf{c}'_\tau, \mathbf{s}'_\tau|\mathbf{s}_t)$ is independent of $\mathbf{s}_t$. The result follows.

References

Abel, Andrew (1990). “Asset prices under habit formation and catching up with the Jones.” *American Economic Review*, **80**, 38-42.

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

Becker, Gary and Murphy, Kevin (1988). “A Theory of Rational Addiction.” *Journal of Political Economy*, **96**, 675-700.

Boldrin, Michele, Lawrence Christiano, and Jonas Fisher (2001). “Habit persistence, asset returns, and the business cycle.” *American Economic Review*, **91**, 149-66.

Bowman, David, Deborah Minehart, and Matthew Rabin (1999). “Loss Aversion in a Consumption-Savings Model.” *Journal of Economic Behavior and Organization*, **38**, 155-178.

Boyd, Norman, Heather J. Sutherland„ Karen Z. Heasman, David L. Tritcher, and Bernard Cummings (1990). “Whose Utilities for Decision Analysis?” *Medical Decision Making*, **10**, 58 - 67.

Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman (1978). “Lottery winners and accident victims: is happiness relative?” *Journal of Personality and Social Psychology*, **36**, 917-27.

Campbell, John and John Cochrane (1999). “By force of habit: A consumption-based explanation of aggregate stock market behavior.” *Journal of Political Economy*, **107**, 205-51.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O’Donoghue, and Matthew Rabin (2002). “Regulation for Conservatives: Behavioral Economics and the Case for ‘Asymmetric Paternalism’.” *Penn Law Review*, forthcoming.

Camerer, Colin, George Loewenstein, and Martin Weber (1989). “The Curse of Knowledge in Economic Settings: An Experimental Analysis.” *Journal of Political Economy*, **97**, 1232-1254.

Carroll, Christopher, Jody Overland, and David Weil (2000). “Saving and growth with habit formation.” *American Economic Review*, **90**, 341-55.

Cohn, Beruria (1999). “The Lay Theory of Happiness: Illusions and Biases in Judging Others.” Undergraduate Dissertation, Princeton University.

Constantinides, George (1990). “Habit formation: A resolution of the equity premium puzzle.” *Journal of Political Economy*, **98**, 519-43.

DellaVigna, Stefano and Ulrike Malmendier (2001). “Self-Control in the Market: Evidence from the Health Club Industry.” Mimeo, Harvard University.

Duesenberry, James (1949). *Income, Saving, and the Theory of Consumer Behavior.* Harvard University Press.

Fischhoff, Baruch (1975). "Hindsight $\neq$ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty." *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 288-299.

Frank, Robert (1999). *Luxury Fever.* New York: Free Press.

Frederick, Shane and GeorgeLoewenstein (1999). "Hedonic Adaptation," in Daniel Kahneman, Edward Diener, and Norbert Schwarz, eds., *Well-Being: The Foundations of Hedonic Psychology.* New York: Russell Sage Foundation Press, 302-329.

Fuhrer, Jeffrey (2000). "Habit formation in consumption and its implications for monetary-policy models." *American Economic Review*, **90**, 367-90.

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

Gilbert, Daniel T., Michael J. Gill, and Timothy D. Wilson (1998). "How do we know what we will like? The informational basis of affective forecasting." Working paper, Harvard University Department of Psychology.

Gilbert, Daniel T., Elizabeth C. Pinel, Timothy D. Wilson, Stephen J. Blumberg, and Thalia P. Wheatley (1998). "Immune neglect: A source of durability bias in affective forecasting." *Journal of Personality and Social Psychology*, **75**(3), 617-638.

Giordano, Louis A., Warren K. Bickel, George Loewenstein, Eric A. Jacobs, Gary J. Badger, and Lisa A. Marsch (2001). "Mild opioid deprivation and delay to consequences affects how opioid-dependent outpatients value an extra maintenance dose of buprenorphine." Substance Abuse Treatment Center, Psychiatry Department, University of Vermont, Burlington.

Helson, Harry (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper and Row.

Herrnstein, Richard, George Loewenstein, Drazen Prelec, and William Vaughan, Jr. (1993). "Utility Maximization and Melioration: Internalities in Individual Choice." *Journal of Behavioral Decision Making*, **6**, 149-185.

Jepson, Christopher, George Loewenstein, and Peter Ubel (2001). "Actual Versus Estimated Differences in Quality of Life Before and After Renal Transplant." Mimeo, Department of Social and Decision Sciences, Carnegie Mellon University.

Jermann, Urban (1998). "Asset pricing in production economies." *Journal of Monetary Economics*, **41**, 257-75.

Kahneman, Daniel (1991). "Judgement and Decision-Making: A Personal View." *Psychological Science*, **2**(3), 142-145.

Kahneman, Daniel (1994). "New challenges to the rationality assumption." *Journal of Institutional and Theoretical Economics*, **150**, 18-36.

Kahneman, Daniel (2000). "Evaluation by Moments: Past and Future," in Daniel Kahneman and Amos Tversky, eds., *Choices, Values, and Frames.* Cambridge University Press and Russell Sage Foundation, 693-708.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1991). "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies." *Journal of Economic Perspectives*, **5**(Winter), 193-206.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, **47**, 263-291.

Kridel, Donald J., Dale E. Lehman, and Dennis L. Weisman (1993). "Option Value, Telecommunications Demand, and Policy." *Information Economics and Policy*, **5**, 125-144.

Laibson, David (1994). "Essays on Hyperbolic Discounting." Dissertation, MIT Department of Economics.

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

Laibson, David, Andrea Repetto, and Jeremy Tobacman (1998). "Self-control and saving for Retirement." *Brookings Papers on Economic Activity*, **1**, 91-196.

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

Loewenstein, George (1999). "A visceral account of addiction," in Jon Elster and Ole-Jørgen Skog, eds., *Getting Hooked: Rationality and Addiction.* Cambridge, England: Cambridge University Press, 235-264.

Loewenstein, George and Daniel Adler (1995). "A Bias in the Prediction of Tastes." *Economic Journal*, **105**, 929-37.

Loewenstein, George and Erik Angner (forthcoming). "Predicting and Honoring Changing Preferences," in George Loewenstein, Daniel Read, and Roy Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice.* New York: Russell Sage Foundation Press.

Loewenstein, George and Shane Frederick (1997). "Predicting Reactions to Environmental Change," in Max Bazerman, David Messinck, Ann Tenbrunsel and Kimberly Wade-Benzoni, eds., *Environment, Ethics, and Behavior.* San Francisco: New Lexington Press, 52-72.

Loewenstein, George, Daniel Nagin, and Raymond Paternoster (1997). "The effect of sexual arousal on predictions of sexual forcefulness." *Journal of Crime and Delinquency*, **34**, 443-73.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2000). "Projection Bias in Predicting Future Utility." U.C. Berkeley Department of Economics Working Paper E00-284.

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: Russell Sage Foundation Press, 85-105.

Muellbauer, John (1988). "Habits, Rationality and Myopia in the Life Cycle Consumption Function." *Annales d'Economie et de Statistique*, **9**, 47-70.

Nisbett, Richard E. and David E. Kanouse (1968). "Obesity, hunger, and supermarket shopping behavior." *Proceedings of the Annual Convention of the American Psychological Association*, **3**, 683-84.

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

O'Donoghue, Ted and Matthew Rabin (1999*b*), "Incentives for Procrastinators." *Quarterly Journal of Economics*, **114**(3), 769-816.

Pollak, Robert A. (1970). "Habit Formation and Dynamic Demand Functions." *Journal of Political Economy*, **78**, 745-763.

Prelec, Drazen and George Loewenstein (1998). "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science*, **17**, 4-28.

Read, Drazen and George Loewenstein (1999). "Enduring pain for money: decisions based on the perception of memory of pain." *Journal of Behavioral Decision Making*, **12**(1), 1-17.

Read, Daniel, and Barbara van Leeuwen (1998). "Predicting hunger. The effects of appetite and delay on choice." *Organizational Behavior and Human Decision Processes*, **76**, 189-205.

Ross, Lee, David Greene, and Pamela House (1977). "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology*, **13**, 279-301.

Ryder, Harl and Geoffrey Heal (1973). "Optimal Growth with Intertemporally Dependent Preferences." *Review of Economic Studies*, **40**, 1-33.

Sackett, David L. and George W. Torrance (1978). "The Utility of Different Health States as Perceived by the General Public." *Journal of Chronic Diseases*, **31**, 697 - 704.

Schulz, Richard and Susan Decker (1985). "Long-term adjustment to physical disability: The role of social support, perceived control, and self-blame." *Journal of Personality and Social Psychology*, **48**(5), 1162-1172.

Scitovsky, Tibor (1976). *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction.* Oxford University Press.

Sieff, Elaine, Robyn Dawes, and George Loewenstein (1999). "Anticipated versus actual responses to HIV test results." *American Journal of Psychology*, **112**(2), 297-311.

Slovic, Paul (2001). "Do cigarette smokers know the risks?" in Paul Slovic, ed., *Smoking: Risk, Perception and Policy.* Thousand Oaks, CA: Sage Publications.

Smith, Adam (2002). *The Theory of Moral Sentiments.* Cambridge University Press (edited by Knud Haakonssen).

Starmer, Chris (2000). "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk." *Journal of Economic Literature*, **38**, 332-82.

Strahilevitz, Michal and George Loewenstein (1998). "The effect of ownership history on the valuation of objects." *Journal of Consumer Research*, **25**, 276-289.

Thaler, Richard (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, **1**, 39-60.

Tversky, Amos and Daniel Kahneman (1991). "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, **106**(4), 1039-61.

Tyc, Vida L. (1992). "Psychosocial adaptation of children and adolescents with limb deficiencies: A review." *Clinical Psychology Review*, **12**, 275-291.

Ubel, Peter, George Loewenstein, Christopher Jepson, Tara Mohr, and Rachel Markowitz (2002). "Quality of life ratings of unfamiliar disabilities: the contribution of a focusing illusion and failure to predict adaptation." Mimeo, University of Michigan Medical Center.

Van Boven, Leaf and George Loewenstein (forthcoming). "Projection of Visceral Feelings." *Personality and Social Psychology Bulletin*.

Van Boven, Leaf, David Dunning, and George Loewenstein (2000). "Egocentric Empathy Gaps Between Owners and Buyers: Misperceptions of the Endowment Effect." *Journal of Personality and Social Psychology*, **79**, 66-76.

Van Boven, Leaf, George Loewenstein, and David Dunning (forthcoming). "Mispredicting the Endowment Effect: Underestimation of Owners' Selling Prices by Buyer's Agents." *Journal of Economic Behavior and Organization*.

Weinstein, Neil (1982). "Community Noise Problems: Evidence Against Adaptation." *Journal of Environmental Psychology*, **2**, 87-97.

Wortman, Camille and Roxane Silver (1987). "Coping with irrevocable loss," in Gary R. VandenBos and Brenda K. Bryant, eds., *Cataclysms, crises and catastrophes: Psychology in action.* Washington, D.C.: American Psychological Association, 189-235.