

Risk, Delay, and Convex Self-Control Costs[†]

By DREW FUDENBERG AND DAVID K. LEVINE*

We develop a dual-self model of self-control that is compatible with modern dynamic macroeconomic theory and evidence. We show that a convex cost of self-control explains a wide range of behavioral anomalies concerning risk, including the Allais paradox, and also explains the observed interaction between risk and delay. We calibrate the model to obtain a quantitative fit. We find that most of the data can be explained with subjective interest rates in the range of 1–7 percent, short-run relative risk aversion of about two, and a time horizon of one day for the short-run self. (JEL D11, D44, D81)

The dual-self model supposes that decision-making involves a conflict between the desire for immediate and safe gratification on the one hand, and riskier long term benefits on the other. This conflict is resolved through a decision criterion that puts weights on both desires. A key element of the model is that this weighting is endogenous, and that changes in constraints that have the same effect on the weights are predicted to lead to the same decisions. The model is consistent with “paradoxes” such as small stakes risk aversion, the common-consequence and common-ratio versions of the Allais paradox, and preference reversals for delayed rewards. It also is consistent with lesser known experimental puzzles, including the effect of reducing the probability of reward on preference reversals induced by delayed rewards, the consequence of delaying rewards for Allais-type preference reversals, preferences for randomization, and reversals due to cognitive load.

In Fudenberg and Levine (2006) we used evidence about cognitive load¹ to argue that the cost of self-control is convex and not linear.² In this paper our first goal is to

*Fudenberg: Department of Economics, Harvard University, 1805 Cambridge Street, Cambridge, MA 02138 (e-mail: dfudenberg@harvard.edu); Levine: Department of Economics, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130 (e-mail: david@dklevine.com). We thank Daniel Benjamin and Jesse Shapiro for helpful comments and a very careful reading of an early draft, and Eduardo Azevedo, Tao Jin, Yuichiro Kamada, and Lauren Merrill for exceptional research assistance. We are also grateful to Juan Carillo, Ed Glaeser, Glenn Harrison, Yoram Halevy, John Kagel, Stefan Krassa, Franco Peracchi, Drazen Prelec, and seminar participants at Harvard, Ohio State, the University of Illinois Champaign Urbana, and the 2006 CEPR conference on Behavioral Economics. NSF grants SES-0426199, SES-0646816, SES-0851315, and SES 0951462 provided financial assistance.

[†]To comment on this article in the online discussion forum, or to view additional materials, visit the article page at <http://www.aeaweb.org/articles.php?doi = 10.1257/mic.3.3.34>.

¹The work of Roy F. Baumeister and collaborators (for example, Mark Muraven, Dianne M. Tice, and Roy F. Baumeister 1998; Muraven and Baumeister 2000; Matthew T. Gailliot et al. 2007) argues that self-control is a limited resource, moreover one that may be measured by blood glucose levels. The stylized fact that people often reward themselves in one domain (for example, food) when exerting more self-control in another (for example, work) has the same implication. This is backed up by evidence from Baba Shiv and Alexander Fedorikhin (1999) and Andrew Ward and Traci Mann (2000) showing that agents under cognitive load exercise less self-control, for example, by eating more desserts. The first two observations fit naturally with the idea that a common “self-control function” controls many nearly simultaneous choices. The third fits naturally with the hypothesis that self-control and some other forms of mental activity draw on related mental systems or resources.

²We also pointed out that convex costs violate the independence axiom of expected utility theory, but did not

explore the implications of convex control costs and to show that they help explain a number of otherwise puzzling experimental “reversals” involving uncertain pay-offs. These kinds of preference reversals have not previously been examined in the self-control literature.³ Our second goal is to explore the quantitative predictions of the model. To do this, we extend the simpler “nightclub” dual-self model of the earlier paper by introducing an additional choice of consumption commitment. A key motivation for the paper is the idea that a model is more useful if its parameters can be held constant across a wider range of settings. There are several competing explanations for each of the facts that our model fits; our quantitative analysis uses a single model with a stable set of parameters to explain many facts.

A central feature of the model is that agents use cash on hand or mental accounting as a commitment device, so that on the margin they will consume all of any small unexpected winnings. However, when agents win large amounts, they choose to exercise self-control and save some of their winnings. The resulting intertemporal smoothing makes the agents less risk averse, so that they are less risk averse to large gambles than to small ones.

In the first part of the paper, we derive new qualitative predictions using our (2006) model. One of these is the interaction of uncertainty and delay: We noted in our previous paper that the dual-self model explains the classic time-inconsistent preferences that have been used to motivate the assumption of quasi-hyperbolic discounting. When there are increasing marginal costs of self-control, there is an additional implication: Preference reversal is less likely when the probability of rewards is smaller. This prediction is borne out in the data of Gideon Keren and Peter Roelofsma (1995) which we learned about after writing the first paper.

We then explain how a convex cost of self-control can explain both the classic Allais paradox (where the independence axiom is violated with respect to mixing in a common consequence) and the “common ratio” version of the paradox. In the Allais paradox there are two scenarios, each involving two options. Under expected utility theory, the same option must be chosen in each scenario, but in practice people choose different options in the two scenarios. A key element of the paradox is that one of the scenarios involves a much smaller probability of winning a prize. That means that there is less temptation to the short-run self. The reason that our model predicts the Allais paradox is that the convexity of the cost function leads to a particular sort of violation of the independence axiom: Agents should be “more rational” about choices that are likely to be payoff-irrelevant. This is exactly the nature of the violation of the independence axiom in the Allais paradox. Importantly, our theory does not explain the other, rarely observed, way that choice in the Allais paradox could violate the independence axiom, namely where the choices in each of the two scenarios are reversed.

explore what sorts of violations occur. In addition we showed how convex cost could lead to violations of the weak axiom of revealed preference, a topic we do not examine here.

³ Jess Benhabib and Alberto Bisin (2005), B. Douglas Bernheim and Antonio Rangel (2004), Isabelle Brocas and Juan D. Carrillo (2008), George Loewenstein and Ted O’Donoghue (2004), and Emre Ozdenoren, Steven W. Salant, and Dan Silverman (forthcoming) present dual-self models, but they do not discuss risk aversion, cognitive load, or the possibility of convex costs of self-control.

An additional prediction of our model is that the paradoxical choices in Allais paradox experiments should disappear when the payoff is outside the time horizon of the short-run self. The data is qualitatively consistent with this but less stark: increasing delay gradually reduces the fraction of paradoxical choices.⁴

In the second part of the paper we turn to a quantitative analysis. Our procedure is to first find a set of sensible values of the key parameters, namely the subjective interest rate, income, the degree of short-term risk aversion, and the time horizon of the short-run self. We focus on the model's ability to fit the behavior of the median individual, although we examine to a limited extent other people such as individuals who are more risk averse than the median.⁵ We then investigate how well we can explain the paradoxes using the calibrated parameter values and the dual-self model. To what extent can the same set of parameter values simultaneously explain the behavior of all the median individuals and be consistent with the calibration?

In the quantitative analysis we use an extension of the basic model that lets it better conform to the data. Specifically, we adopt a nested CES/logarithmic specification that allows a greater level of risk aversion in the laboratory,⁶ and combine this with a consumption commitment technology. We then give a quantitative analysis of three different paradoxes. We examine Matthew Rabin (2000) paradox data from Charles A. Holt and Susan K. Laury (2002), both the Daniel Kahneman and Amos Tversky (1979) and Allais versions of the Allais paradox, and the finding of Daniel J. Benjamin, Sebastian A. Brown, and Jesse M. Shapiro (2006) that increased cognitive load increases risk aversion.

We find that we can explain most of the data if we assume an annual subjective interest rate in the range of 1–7 percent, a short-run coefficient of relative risk aversion of about two, and a daily time horizon for the short-run self. We find that the Rabin paradox is relatively insensitive to the exact parameters assumed; the Allais paradox is sensitive to choosing a plausible level of risk aversion; and the cognitive load data is very sensitive to the exact parameter values chosen.

⁴ See Keren and Roelofsma (1995) and Bethany J. Weber and Gretchen B. Chapman (2005) for the effect of delay on the common-consequence Allais paradox and Manel Baucells and Franz H. Heukamp (2010a) for the effect of delay on the common-ratio version. In Fudenberg and Levine (2010) we extend the dual-self model to account for the gradual effect of delay on both the frequency of paradoxical choices and also on the implicit marginal subjective interest rate, as in the work surveyed by Shane Frederick, George Loewenstein, and Ted O'Donoghue (2002).

⁵ Of course there is no a priori theoretical reason for the subject pools in the different experiments to have the same preferences. The justifications for our approach are that the theory is more powerful if the parameters are consistent across pools, and that there is little point of doing experimental or empirical work if one takes the view that each small group of people may have preferences totally unlike those of any other group.

⁶ Our earlier model can explain the examples in Rabin (2000), but those examples (such as rejecting a bet that had equal probability of winning \$105 or losing \$100) understate the degree of risk aversion in small-stakes experiments, and fitting our earlier model to these small gambles requires parameter values that conflict both with intuition and with other data. Since the first version of this paper was written, James C. Cox et al. (2007) conducted a series of experiments to test various utility theories using relatively high stakes. They also observe that the simple logarithmic model is inconsistent with observed risk aversion, and they argue that the simple linear-logarithmic self-control model does not plausibly explain their data. We will be interested to see whether their data is consistent with the more complex model developed here.

I. The Base Model

To begin we review the base model from our 2006 paper. We consider an infinite-lived consumer making a savings decision. Each period $t = 1, 2, \dots$ is divided into two sub-periods, the “bank” sub-period and the “nightclub” sub-period; the consumer discounts utility between periods using discount factor $\delta \in (0, 1)$, but no discounting occurs between the sub-periods. Wealth at the beginning of the bank sub-period is denoted by w_t . During the “bank” sub-period, consumption is not possible, and wealth is divided between savings s_t , which remains in the bank, pocket cash x_t , which is carried to the nightclub. In the nightclub consumption $c_t \in [0, x_t]$ is determined, with $x_t - c_t$ returned to the bank at the end of the period. Wealth next period is just $w_{t+1} = R(s_t + x_t - c_t)$, where $R \geq 0$ is the risk-free interest rate. For simplicity money returned to the bank bears the same rate of interest as money left in the bank.

Our underlying rationale is that perfect capital markets are available at the bank, so we can capitalize all of the consumer’s future income into her initial wealth w_1 . The only constraint at the bank then is that wealth w_t must be non-negative. By way of contrast, capital markets are not available at the nightclub, and no income is received there, so the only choice at the nightclub is how much pocket cash to spend.

The utility of the short-run self of consuming c_t in period t is $u(c_t)$. We assume this is strictly differentiable concave⁷ and satisfies the Inada condition that $\lim_{c_t \rightarrow 0} u'(c_t) = \infty$. The long-run self maximizes the expected discounted present value of the utility of the short-run selves, subject to a cost of self-control. This self-control cost depends on the resources the short-run self perceives as available to himself, which in turn determine a *temptation utility* for the short-run self, representing the utility the short-run self perceives as available if no self-control is used. Denote this temptation utility by \bar{u}_t . The actual realized utility that the long-run self allows the short-run self is u_t , and the cost of self-control is $g(\bar{u}_t - u_t)$, where $g(0) = 0$ and g is continuously differentiable, non-decreasing, and at least weakly convex. One key idea here, as in Faruk Gul and Wolfgang Piesendorfer (2001), is that the cost of self-control depends on the difference between the utility by which the short-run self is tempted, \bar{u}_t , and the utility the short-run self is allowed, u_t . A second key idea is that we do not require g to be linear, and several of theoretical and empirical findings hinge on the idea that g is typically convex, as we argued in Fudenberg and Levine (2006).⁸

In the bank no consumption is possible, and so there is no temptation for the short-run self. In the nightclub the short-run self cannot borrow, and wishes to spend all of the available pocket cash x_t on consumption. Hence $\bar{u}_t = u(x_t)$.

The problem faced by the long-run self is to choose pocket cash and consumption to maximize the present value using the discount factor δ of short-run self utility net of the cost of self-control. The objective function of the long-run self is

$$(1) \quad U_{RF} = \sum_{t=1}^{\infty} \delta^{t-1} [u(c_t) - g(u(x_t) - u(c_t))]$$

⁷ That is, it is twice continuously differentiable on the interior, and the second derivative is strictly negative there.

⁸ Jawwad Noor and Norio Takeoka (2010a, 2010b) also emphasize the importance of a convex cost of self-control.

which is to be maximized with respect to $c_t \geq 0$, $x_t \geq 0$ subject to w_1 given, $w_{t+1} = R(s_t + x_t - c_t)$, $s_t + x_t \leq w_t$ and $w_t \geq 0$. Notice that this is a simple optimization problem with no uncertainty and perfect foresight.

A crucial aspect of the model is the “pocket cash” x_t that serves to ration consumption and thus reduces the temptation to the short-run self. This commitment lasts exactly one period. Given the one-period horizon of the short-run self, a longer commitment would have no benefits but could reduce flexibility.⁹ When there is no uncertainty a one-period commitment allows the long-run self to attain the same optimum as if there was no cost to self-control. The reason for this is that it is optimal to give the short-run self exactly the amount to be spent at the nightclub, and so avoid temptation and self-control cost entirely. Underlying this is the assumption that the decision about how much pocket cash to allocate to the short-run self is taken when there are no tempting consumption possibilities. For that reason, the solution to the perfect-foresight problem is to choose $c_t = x_t$. In other words, cash x_t is chosen to equal the optimal consumption for an agent without self-control costs. The agent then spends all pocket cash at the nightclub, and so incurs no self-control cost there. Since $c_t = x_t$, the utility of the short-run self is $u(x_t)$, and as there is no self-control cost, this boils down to maximizing

$$\sum_{t=1}^{\infty} \delta^{t-1} u(x_t)$$

subject to the budget constraint $w_{t+1} = R(w_t - x_t)$. Denote the solution to this problem as \hat{x}_t , and the corresponding value function $V(w_t)$.

In our earlier paper, the notion of a bank and pocket cash were taken literally. In practice there are many strategies that individuals use to reduce the temptation for impulsive expenditures. The view we take here is that pocket cash is determined by mental accounting of the type discussed by Richard Thaler (1980), and not necessarily by physically isolating money in a bank. In other words, x_t should not be viewed as the literal amount of money the short-run self has in their wallet or the amount available including cash cards and so forth, but should be viewed as the amount of resources that the short-run self feels entitled to use. The strategies individuals use for this type of commitment can be varied. For example some people may choose to carry only a limited amount of cash and no credit cards, which would implement the consumption commitment but at a cost in terms of trips to the ATM, risk of theft, and so forth. Others may be able to use mental commitments and prefer to do so, for example with a plan such as “you may spend \$100 at the nightclub, but no more.”

The key point is that since many people implement these types of controls on impulsive behavior through mental rather than physical rationing we cannot identify “pocket cash” x_t with the physical amount of money carried in an individual’s wallet. However, although mental accounting is not directly observable, the theory implies that x_t is the expenditure during the horizon of the short-run self, and so it

⁹ If the short-run self is tempted by future consumption as well as by consumption in the current period, as in Fudenberg and Levine (2010), Noor (2007; 2011), and Noor and Takeoka (2010b), there is then a trade-off involved in determining the optimal commitment length; we hope to explore this in future work.

is computable directly from the underlying preference parameters. Moreover, since the horizon of the short-run self is on the order of a day to a week, there is an external check on the plausibility of the computed values. To the extent that the calibrated parameters are stable across situations, the implicit pocket cash will be as well.

II. Risky Drinking: Nightclubs and Lotteries

Suppose in period 1 (only) that when the agent arrives at the nightclub of her choice, she has the choice between two lotteries, S and L , with intertemporal returns $\tilde{z}_t^S, \tilde{z}_t^L |_{t=1}^\infty$, and with the entire time stream of the lottery's payout determined in period 1. Note that we use tildes in the notation here to indicate random variables: we will use z_t^S and so forth without tildes for realizations of the random variables. We will adopt the convention that when there is a high self-control cost preferences for the lottery S are stronger (relative to L) than when there is no self-control cost. Intuitively we think of S as better liked by the short-run self and L by the long-run self. In our applications the random receipts will be independent between periods. Initially we will assume that this choice is completely unanticipated—that is, it has prior probability zero. What is the optimal choice of lottery given x_1 ? For simplicity, we assume throughout that the agent does not expect to have more choices between lotteries at nightclubs after period 1. We should emphasize that this is for convenience only: the overall savings and utility decision will not change significantly provided that the probability of getting future choice opportunities is small.¹⁰

The returns $z_t^S, z_t^L |_{t=1}^\infty$ may be positive or negative, but we suppose that the largest possible loss in period 1 is less than the agent's pocket cash. There are a number of different ways that the dual-self model can be applied to this setting, depending on the timing and “temptingness” of the choice of lottery and spending of its proceeds. In this paper, we assume that in the nightclub the agent simultaneously decides which lottery to pick and how to spend in the first period for each possible realization of the lottery.

Let V denote the maximized value of LR's discounted expected payoff, as a function of wealth. (From the stationarity of the problem this value does not depend on t .) Then the decision problem in period 1 is to choose consumption $\tilde{c}_1^j(z_1^j)$ contingent on the realization of the lottery j and either $j = S$ or $j = L$ to maximize

$$\begin{aligned} Eu(\tilde{c}_1^j) - g(\max\{Eu(x_1 + \tilde{z}_1^S), Eu(x_1 + \tilde{z}_1^L)\} - Eu(\tilde{c}_1^j)) \\ + \delta EV(R(w_1 + \tilde{z}_1^j - \tilde{c}_1^j) + \sum_{t=2}^\infty R^{2-t} \tilde{z}_t^j). \end{aligned}$$

That is, if no self-control is exercised the short-run self will obtain $\max\{Eu(x_1 + \tilde{z}_1^S), Eu(x_1 + \tilde{z}_1^L)\}$, and so this temptation utility is used to determine the cost of self-control for the long-run self. Note that this formulation applies because the lottery's first period return occurs at the night club and can be spent then and because the lottery and consumption decision are made simultaneously. If the consumption were to

¹⁰ This was demonstrated formally in an earlier working paper version (Fudenberg and Levine 2010).

occur in a “subsequent” period then it would not be tempting for the short-run self, and there would be no temptation in either period.¹¹ Note also that because g need not be linear, $g(Eu)$ need not equal $E(g(u))$; when g is linear the equality holds and the objective function is an expected utility.

In what follows, it will be useful to refer to the agent’s wealth following a particular realization of $\tilde{z}_t^S, \tilde{z}_t^L|_{t=1}^\infty$, which we denote

$$w_2^j(c_1^j, z_1^j|_{t=1}^\infty) \equiv R(w_1 + z_1^j - c_1^j) + \sum_{t=2}^\infty R^{2-t} z_t^j.$$

From here on, when the $\tilde{z}_t^j|_{t=1}^\infty$ is clear, we will omit it as an explicit argument.

Consider first the problem of determining the optimal consumption in period 1 as a function of the income receipt z_1^j . Let γ_1 denote the marginal cost of self-control in the first period. We let $\hat{c}_1^j(\gamma_1)(z_1^j)$ be the solution to the first-order condition for a maximum for a fixed marginal cost of self-control γ_1 . In this case of linear control cost, the overall objective function is strictly differentiable concave in the scalar choice variable, c_1^j , so this solution exists and is unique. Given a consumption derived from γ_1 in this way we can compute a new marginal cost of self-control

$$\begin{aligned} \hat{\gamma}_1^j(\gamma_1) = & g'(\max\{Eu(x_1 + \tilde{z}_1^S), Eu(x_1 + \tilde{z}_1^L)\}) \\ & - Eu(\min\{\hat{c}_1^j(\gamma_1)(\tilde{z}_1^j), x_1 + \tilde{z}_1^j\}). \end{aligned}$$

Theorem 1 establishes that “the” marginal cost of self-control is the unique fixed point of this map.

THEOREM 1: *For given x_1 and each $j \in \{S, L\}$ there is a unique solution to*

$$\gamma_1^j = \hat{\gamma}_1^j(\gamma_1^j)$$

and this solution together with $\tilde{c}_1^j = \min\{\hat{c}_1^j(\gamma_1^j)(z_1^j), x_1 + z_1^j\}$ and the choice of j that maximizes long-run utility is necessary and sufficient for an optimal solution to the consumer’s choice between lotteries S and L ,

PROOF:

See Appendix A.

The “consumption function” is $\tilde{c}_1^j = \min\{\hat{c}_1^j(\gamma_1^j)(z_1^j), x_1 + z_1^j\}$. Let \hat{z}_1 be the maximal value of z_1^j such that all the winnings are spent, that is, $\hat{c}_1^j(\gamma_1)(\hat{z}_1) \geq x_1 + \hat{z}_1$.¹² The function \hat{c}_1^j is sketched in Figure 1. For $z_1^j < \hat{z}_1$ no self-control is used, and all winnings are spent. Above this level self-control is used, with only a fraction of

¹¹ The model here is very stark, with each short-run self-living for a fixed length of time, so every choice is either “completely tempting” (when it is in the lifetime of the current short-run self) or not tempting at all. One way to relax this dichotomy is to assume that the short-run player has a stochastic lifetime. We pursue this theory in other work, showing that the results here are robust to the exact timing of the choices.

¹² Note that for arbitrary values of x_1 we may have \hat{z}_1 negative, but this will not happen on a perfect-foresight path.

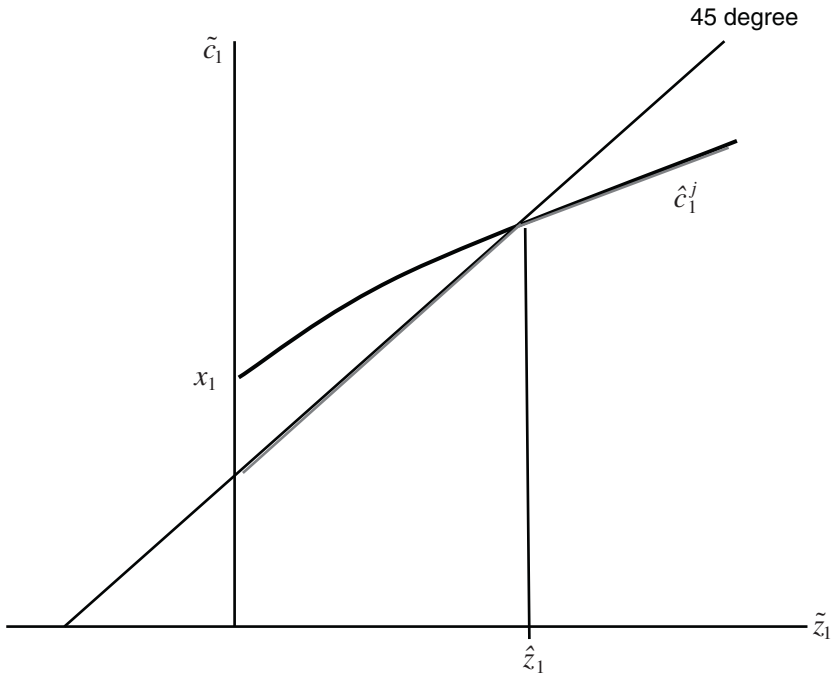


FIGURE 1. THE CONSUMPTION FUNCTION

winnings consumed, and the rest going to savings. The size of \hat{z}_1 depends, among other things, on the agent's cost of self-control; if this is very high then the agent will have a high propensity to consume even from large winnings, while if it is low the agent will exert self-control even over small windfalls. When the agent is very patient in the sense of having a subjective discount factor δ close to one, \hat{c}_1^j is very flat, so that only a tiny fraction of the winnings are consumed immediately when receipts exceed the critical level. Thus when the agent is patient he is almost risk neutral with respect to gambles that are larger than his usual per-period consumption. However the agent is still risk averse to small gambles, as these will not be smoothed but will lead to a one for one change in current consumption.¹³

III. Qualitative Analysis

We now turn to a qualitative analysis of the choice between lotteries. We consider a variety of conceptual experiments involving both one-period and intertemporal choices. As in the previous section we suppose that the various "experimentally induced" choices the agent faces in the nightclub were completely unforeseen; this

¹³ Note that the nondifferentiability of consumption and thus of indirect utility in our model occurs at the self-control threshold, which is strictly positive and is endogenous to the model (but exogenous to the specific gamble). Thus the non-differentiability is different than the kink at zero in loss-aversion models, and also different than kinks at "reference points" that vary with the gamble under consideration.

may be a good description of most experimental studies.¹⁴ The key fact is that the short-run self is less patient than the long-run self and also (because of the pocket cash constraint) more risk averse; we show this formally in Theorem 2. However, the weight given to the preferences of the current short-run self depends on the marginal cost of self-control. When temptation is high, the marginal cost of self-control is high, and a great deal of weight should be given to the impatient risk-averse preferences of the current short-run self. When the temptation is low, the marginal cost of self-control is low and little weight is given to the impatient risk-averse preferences of the short-run self. Temptation depends on expected payoffs, which in turn depend on the stakes, on the discounting of future utilities, and on the probability distribution over outcomes. When changes in timing or probability reduce the temptation, the presence of self-control costs can lead to a “reversal” from an alternative preferred by the short-run self to one less preferred. As we shall see, these sorts of reversals are a fairly common feature in experimental data; our model provides a common coherent explanation of them. Noor and Takeoka (2010b) provide an axiomatic characterization of a related but different form of nonlinearity, and shows how it is connected to Allais-type paradoxes and the interaction of risk and delay.

Risk and Delay.—We noted in our previous paper that the dual-self model explains the kind of time-inconsistent preferences that have been used to motivate the assumption of quasi-hyperbolic discounting. Here we derive an additional implication in the case of increasing marginal costs of self-control: Preference reversal due to present bias is less likely when the probability of rewards is smaller.

Specifically, suppose that unexpectedly in period 1 the short-run self at the nightclub is offered a choice between an amount z_t^S in period t and an amount $z_{t+1}^L = \theta z_t^S$ in period $t + 1$ where $\theta > R$. If period t is in the future, that is, $t > 1$, the short-run self is indifferent and since $\theta > R$ the long-run self strictly prefers to implement the choice of option L .¹⁵

The situation is different if $t = 1$. Notice that for small amounts $z_1^S < \hat{z}_1$, if the amount today is taken it is consumed immediately. Hence the utility gain from choosing S is

$$\begin{aligned} \Delta &= u(x_1 + z_1^S) + \delta EV(w_2) - u(x_1) + g(u(x_1 + z_1^S) - u(x_1)) \\ &\quad - \delta V(w_2 + \theta z_1^S). \end{aligned}$$

Differentiating this with respect to z_1^S we find

$$\frac{d\Delta}{dz_1^S} \Big|_{z_1^S=0} = (1 + g'(0))u'(x_1) - \delta\theta V'(w_2).$$

¹⁴ The results of the analysis are robust to the case in which the choices are foreseen but are thought a priori to have very low probability. This was demonstrated formally in an earlier working paper version.

¹⁵ Because the long-run self has access to a perfect capital market, he acts to maximize lifetime wealth, and the subjective discount factor δ is irrelevant. Notice that while borrowing and lending is possible, it is possible only at the bank, and so does not pose a temptation for the short-run self.

TABLE 1—DYNAMIC PREFERENCE REVERSAL

Scenario		Probability of reward ^a	
		1.0 (60)	0.5 (100)
1	S \$175 now	0.82	0.39
	L \$192 4 weeks	0.18	0.61
2	S \$175 26 weeks	0.37	0.33
	L \$192 30 weeks	0.63	0.67

Note:

^aSample size in parentheses.

The first-order condition for the value function without self-control that defines V implies that $V'(w_2) = (1/R\delta)u'(x_1)$. Thus $d\Delta/dz_1^S = (1 + g'(0) - \theta/R)u'(x_1)$, so if $g'(0) > \theta/R - 1$ then S is the preferred alternative for small z_1^S .

Here is our first example of a “reversal,” in the sense that the willingness to wait in period 1 for higher payoff in period 2 is different than willingness to postpone payoffs from period $t > 1$ to $t + 1$. Notice that this is a first-order reversal in the sense that it arises from the reduction of temptation by changing the timing of the consumption, and this reduction occurs even with linear control costs.

The fact that some subjects exhibit such a reversal has been used to motivate the assumption of quasi-hyperbolic utility. In our model, this reversal arises because at the nightclub the short-run self is rationed by the available cash x_1 . An extra cash payment of z_1 today will cause a temptation to increase spending that is costly for the long-run self to control. By contrast there is no temptation associated with future payoffs, and so there can be “preference reversal” whenever the cost of resisting the short-run temptation is sufficiently high. Data from Keren and Roelofsma (1995) in Table 1, in which the “probability of reward” is equal to 1.0, shows how this happens in practice.¹⁶

Here 82 percent of the population will take the smaller but more immediate reward, when the earlier reward is “now” but only 37 percent will do so when the earlier reward is not for 26 weeks.

Now we turn to the possibility of a second-order reversal due to the change in the marginal cost of self-control when temptation is reduced. Suppose that instead of a certain reward z_t^S, z_{t+1}^L there is only a chance p of getting the reward. When $t > 1$ standard expected utility theory applies, and the fact that rewards have probability less than 1 makes no difference; here L is still strictly preferred as long as $\theta > R$. Consider, however, the utility difference for $t = 1$. Still assuming $z_1^S < \hat{z}_1$ so that $c_1^S = x_1 + z_1^S$ we have

$$\begin{aligned} \Delta(p) = & p[u(c_1^S) + \delta V(R(w_1 + z_1^S - c_1^S))] \\ & - p[u(x_1) + \delta V(R(w_1 - c_1^L) + \theta z_1^S)] + g(p[u(x_1 + z_1^S) - u(x_1)]). \end{aligned}$$

¹⁶ This experimental result is confirmed by Weber and Chapman (2005), and discussed in Yoram Halevy (2008), who proposed an objective function that is consistent with these choices. Note that the experiment was in Dutch florins. We converted from Dutch florins to US dollars using an exchange rate typical of the early 1990s of 1.75 florin per dollar.

To evaluate $\Delta(p)$ for p near 1 we will use Taylor's theorem so we first compute $d\Delta/dp$:

$$\begin{aligned} \frac{d\Delta}{dp} \Big|_{p=1} &= \Delta(1) + g'(u(x_1 + z_1^S) - u(x_1)) [u(x_1 + z_1^S) - u(x_1)] \\ &\quad - g([u(x_1 + z_1^S) - u(x_1)]) \\ &> \Delta(1), \end{aligned}$$

where the final inequality holds strictly if $g'(\cdot)$ is strictly increasing. In particular, if $\Delta(1) = 0$ so there is exact indifference when $p = 1$, then a small decrease in p will imply $\Delta(p) < 0$, so that L will be strictly preferred. By continuity, if $\Delta(1)$ is slightly positive so that S is preferred, then there will be $p < 1$ such that L is preferred. Or put differently, as p is reduced, behavior at $t = 1$ comes to resemble that for $t > 1$. This dependence of the choices on the probability of reward is not consistent with quasi-hyperbolic preferences (as in David Laibson 1997) or with the version of the independence axiom (for choices over menus) imposed as an axiom by Gul and Pesendorfer (2001) and Eddie Dekel, Barton Lipman, and Aldo Rustichini (2009). It is, however, in accord with the Keren and Roelofsma (1995) data in Table 1, where when p is reduced from one to 0.5 in both scenario 1 and scenario 2 the probability of choosing the early reward S is quite similar, 0.39 and 0.33 respectively, and also very similar to the probability of choosing the early reward when $p = 1$ and $t > 1$, which is 0.37.

Risk Aversion in the Short-Run and the Long-Run.—We now wish to focus on reversals that arise from risky prospects. To do this we need to establish the sense in which the short-run self is more risk averse than the long-run self. We do this by establishing a basic relationship between long-run risk aversion as measured by the value function and short-term risk aversion as measured by the utility function

THEOREM 2: *Let $\xi \equiv \sup_{c_1} (-c_1 u''(c_1)/u'(c_1)) / \inf_{c_1} (-c_1 u''(c_1)/u'(c_1))$. Then*

$$-\frac{V''(w_1)}{V'(w_1)} \leq -\xi \frac{c_1}{w_1} \frac{u''(c_1)}{u'(c_1)}.$$

PROOF:

See Appendix I.

This says that the “long-run” absolute risk aversion as measured by the value function is much less than “short-run” absolute risk aversion as measured by the period utility function provided that the ratio of consumption to wealth is small. In this model period length is the time-horizon of the short-run self, which is on the order of a day. The ratio of daily consumption to lifetime wealth is quite small. This means that consumption that is spread over the lifetime represents many magnitudes less of risk than consumption that is taken in the initial period only. In this model, with positive self-control, we have already observed that unanticipated gains less than \hat{z}_1 are all consumed immediately, so they are evaluated using u . Larger or

anticipated gains are spread over the entire lifetime, and are evaluated with the less risk-averse utility for lifetime wealth, which explain the discrepancy between the very high small-stakes risk aversion in the lab and less risk averse long-run decisions that is known as the “Rabin paradox.” For example, using observed absolute risk aversion in the laboratory, if we fit a constant relative risk aversion utility function with respect to lifetime wealth, the estimated coefficient of relative risk aversion is on the order of 30,000,¹⁷ while stock market behavior reveals a coefficient of absolute risk aversion corresponding to relative risk aversion of no more than 48,¹⁸ while macroeconomic studies often take the coefficient of relative risk aversion to be one (that is, logarithmic utility). The huge difference between the lab-based numbers and the investment-based estimates are what we would expect if the laboratory winnings are smaller than \hat{z}_1 , and stock market gains and losses much larger; we explore this quantitatively in later sections. This differential between risk aversion for small and large gains also plays a key role in our analysis of reversals such as the Allais and common ratio paradoxes.

The Allais and Common-Ratio Paradoxes.—We turn now to the classical reversals of the Allais and Common-Ratio paradoxes. In both cases the reversal involves switching from a less risky to a more risky alternative when the chances of winning a prize are reduced. These scenarios are period one gambles, so $\tilde{z}_t^j = 0$, $t > 1$. For simplicity we consider three possible outcomes $Z_b < Z_m < Z_g$ where Z_b is taken to be zero.

We consider two different scenarios. In the base scenario the probabilities of the outcomes $i \in \{b, m, g\}$ are p_i^j under the alternatives $j = S, L$. In the second scenario these probabilities are \bar{p}_i^j .

In the Allais paradox the probability of the good outcome is the same in both scenarios, that is, $\bar{p}_g^j = p_g^j$, while the probability of the middle outcome is reduced by the same amount for both actions in moving from the first scenario to the second, so $\bar{p}_m^j = p_m^j - P$. For example, in the classical Allais paradox, the rewards are $Z_b = 0$, $Z_m = 1,000,000$, $Z_g = 5,000,000$, while the base probabilities are $p^S = (0, 1, 0)$, meaning the middle outcome is certain, and $p^L = (0.01, 0.89, 0.10)$. The second scenario is defined by $P = 0.89$, meaning that $\bar{p}^S = (0.89, 0.11, 0)$ and $\bar{p}^L = (0.9, 0, 0.10)$. An expected utility maximizer must make the same choice in each scenario, but it is often observed that when asked about these hypothetical choices, many subjects prefer S in the base scenario and L in the second scenario.

In the common-ratio paradox, the probability of both the middle and good outcome are reduced by a common factor α , so that $\bar{p}_m^j = \alpha p_m^j$, $\bar{p}_g^j = \alpha p_g^j$. In Baucells and Heukamp (2010a), for example, $Z_b = 0$, $Z_m = 9$, $Z_g = 12$. The base probabilities are $p^S = (0, 1, 0)$, $p^L = (0.2, 0, 0.8)$, and $\alpha = 0.1$ so that $\bar{p}^S = (0.9, 0.1, 0)$, $\bar{p}^L = (0.92, 0, 0.08)$. As with the Allais paradox, an expected utility maximizer makes the same choice in each scenario, yet in practice many choose S in the base scenario and switch to L in the second scenario. There are many experiments showing the

¹⁷ For example if we use the experimental data from Peter Boessarts, Charles Plott, and William R. Zame (2007) and a lifetime wealth of \$350,000.

¹⁸ Rajnish Mehra (2003).

TABLE 2—COMMON RATIO PARADOX

<i>S</i> 1.00 chance of 9 euros	0.58
<i>L</i> 0.80 chance of 12 euros	0.42
<i>S</i> 0.10 chance of 9 euros	0.22
<i>L</i> 0.08 chance of 12 euros	0.78

common ratio reversal. One such is Baucells and Heukamp (2010a) whose data is especially useful for us because they examine the temporal dimension of the reversal which we examine below.¹⁹ Their data in Table 2 show that only 22 percent choose the safer choice *S* in the second scenario, while the substantially greater fraction 58 percent make that choice in the base scenario.²⁰

Notice that in the base scenario *S* has a lower mean than *L* and *S* is less risky. Consequently if *L* is preferred by *u* to *S* then it is also preferred by any less risk averse utility function,²¹ and by Theorem 2 if c_1/w_1 is small then *V* is less risk averse than *u*. Since both *u* and *V* are expected utility functions, and since the common ratio scenarios are constructed so that for any expected utility function preferences must be the same in the two scenarios, if *L* is preferred to *S* by *u* and *V* in the base scenario it is also preferred by both in the second scenario. Consequently no reversal will occur.

The interesting case, therefore, is the case in which the short-run self prefers *S*, that is, $Eu(x_1 + \tilde{z}_1^S) \geq Eu(x_1 + \tilde{z}_1^L)$, which we now assume. The idea is that in moving from the base scenario to the second scenario the overall probability of winning a prize is reduced, so the temptation and marginal cost of self-control are reduced, leading to a reversal from *S* to *L*.

To demonstrate this formally we examine the case where $p_i^L = p_i^S + \lambda P_i$ and $\bar{p}_i^L = \bar{p}_i^S + \lambda \bar{P}_i$ for $i = b, m, g$, and $P_b = -P_m - P_g$, $\bar{P}_b = -\bar{P}_m - \bar{P}_g$ for small positive λ . In the Allais paradox $\bar{p}_g^j = p_g^j$, and $\bar{p}_m^j = p_m^j - P$. In other words $\bar{P}_i = P_i$. In the common ratio paradox $\bar{p}_m^j = \alpha p_m^j$, $\bar{p}_g^j = \alpha p_g^j$ so $\bar{P}_i = \alpha P_i$. Note that while the equations for \bar{p}_i^j are different for the Allais and common ratio paradoxes, both are special cases of what we may think of as the “generalized” common ratio paradox $\bar{P}_i = \alpha P_i$. For the Allais paradox, we have $\alpha = 1$, for the common ratio paradox $\alpha < 1$. We will now analyze the consequences of assuming $\bar{P}_i = \alpha P_i$.

To determine the optimal choice, define Δ to be the utility advantage of *S* in the base scenario:

$$\begin{aligned} \Delta = & \{Eu(\tilde{c}_1^S) - g(Eu(x_1 + \tilde{z}_1^S) - Eu(\tilde{c}_1^S)) + \delta EV(w_2(\tilde{c}_1^S))\} \\ & - \{Eu(\tilde{c}_1^L) - g(Eu(x_1 + \tilde{z}_1^S) - Eu(\tilde{c}_1^L)) + \delta EV(w_2(\tilde{c}_1^L))\}. \end{aligned}$$

¹⁹ While Keren and Roelofsma (1995) find that the immediacy effect is attenuated by uncertainty of reward, and Baucells and Heukamp (2010a) find that the common ratio paradox is attenuated by delay, Weber and Chapman (2005) are unable to find either effect, albeit in a very small sample. They also fail to find any attenuation of the common consequence Allais paradox due to delay. Hence the importance of these effects must await further research.

²⁰ Based on a sample of 221 subjects.

²¹ Recall that *u* is less risk averse than *v* if $v = g(u)$ where *g* is strictly increasing and weakly concave.

For $\lambda = 0$ the two lotteries are equivalent and so $\Delta(0) = 0$. To differentiate this with respect to λ notice that by the envelope theorem the derivatives with respect to the consumption levels $c_{i_i}^j$ are zero, so we need to consider only the derivative with respect to the probabilities p_i^j in the expectations. From $p_i^L = p_i^S + \lambda P_i$, the derivatives of the differences in probabilities between the two choices are just P_i . Hence we may compute

$$\begin{aligned} \frac{d\Delta}{d\lambda} \Big|_{\lambda=0} &= [1 + g'(Eu(x_1 + \tilde{z}_1^S) - Eu(\tilde{c}_1^S))] \\ &\quad \times [(u(c_{1m}^S) - u(c_{1b}^S))P_m + (u(c_{1g}^S) - u(c_{1b}^S))P_g] \\ &\quad + \delta[(V(w_2(c_{1m}^S)) - V(w_2(c_{1b}^S)))P_m + (V(w_2(c_{1g}^S)) - V(w_2(c_{1b}^S)))P_g]. \end{aligned}$$

If this is positive then for small λ^{22} the optimal choice is L , and if it is negative the optimal choice is S .

When can changing from the base scenario to the second scenario reverse the agent's preferences? Let $\bar{\Delta}$ be the utility advantage of S in the second scenario. If the marginal cost of self-control, namely $\gamma_1 = g'(Eu(x_1 + \tilde{z}_1^S) - Eu(\tilde{c}_1^S))$, is the same in both scenarios then

$$\frac{d\bar{\Delta}}{d\lambda} = \frac{\alpha d\Delta}{d\lambda}$$

so that there can be no sign change and no reversal. If

$$U^* = (u(c_{1m}^S) - u(c_{1b}^S))P_m + (u(c_{1g}^S) - u(c_{1b}^S))P_g$$

and

$$V^* = (V(w_2(c_{1m}^S)) - V(w_2(c_{1b}^S)))P_m + (V(w_2(c_{1g}^S)) - V(w_2(c_{1b}^S)))P_g$$

have the same sign, then the magnitude of γ_1 will not matter, and again there will be no reversal. Since u is more risk averse than V the interesting case then is one in which U^* is negative (S preferred) and V^* is positive (L preferred).

Let us suppose that $Z_b < \hat{z}_1$. Then

$$\gamma_1 = g'((u(x_1 + z_{1m}^S) - Eu(\tilde{c}_{1m}^S))p_m^S + (u(x_1 + z_{1g}^S) - Eu(\tilde{c}_{1g}^S))p_g^S).$$

In both the Allais paradox and the common ratio paradox $\bar{p}_m^S < p_m^S$, $\bar{p}_m^L \leq p_m^L$, so if $Z_m > \hat{z}_1$ and $g'(\cdot)$ is strictly increasing $\bar{\gamma}_1 < \gamma_1$. If $d\Delta/d\lambda = 0$, this implies $d\bar{\Delta}/d\lambda < 0$, so by continuity the same is true if $d\Delta/d\lambda$ is slightly positive. In other words, the dual self model with strictly convex cost of self-control predicts both an Allais and common ratio paradox.

²² The general case of large λ is considered for the Allais paradox below.

TABLE 3—COMMON RATIO WITH DELAY

	Now	3 month delay
<i>S</i> 1.00 chance of 9 euros	0.58	0.43
<i>L</i> 0.80 chance of 12 euros		
<i>S</i> 0.10 chance of 9 euros	0.22	0.21
<i>L</i> 0.08 chance of 12 euros		

Reversal and Delay.—Dual self theory has a second implication for Allais and common ratio paradoxes: if the results of gambles are delayed long-enough that they fall outside the time horizon of the short-run self, the reversals should disappear. Some people find this conclusion to be implausible. There is very little data on what happens to Allais type reversals when there is delay, but Baucells and Heukamp (2010a) have studied the common ratio reversal with and without delay. Their data is summarized in Table 3.

The key fact is that the reversal rate is 0.36 without delay and falls to 0.22 with delay, which is qualitatively as the theory predicts. However, quantitatively fitting this data requires that the subjects who exhibit reversals with the 3-month delay have a “short run” of more than three months, which is not consistent with calibrations of the short run on other experiments. We address this fact, and other evidence that the length of delay has a continuous impact on decisions, in Fudenberg and Levine (2010).

Betweenness.—Another implication of increasing marginal cost of self-control is that it may be desirable to incur some cost of self-control, but not too much. This can result in the choice of a “compromise” action that lies in between an option that involves a high and a low cost of self-control. In particular—a dual-self agent may strictly prefer to randomize as a compromise to achieve and “intermediate” level of self-control cost.

Specifically, suppose that in addition to *S* and *L*, there is a third option *M* consisting of a probability $0 < \lambda < 1$ of *S* and $1 - \lambda$ of *L*. In expected utility theory unless *S* and *L* are indifferent, *M* can never be chosen, and indeed, can never be strictly preferred over both *S* and *L*. Dekel (1986) proposed this “betweenness” property as a way of relaxing the independence axiom. However, Colin F. Camerer and Teck-Hua Ho (1994) summarize considerable past work showing that this axiom is violated by a non-trivial fraction of subjects, and also conducted their own experiments that yielded similar results. For example, 20 percent of their subjects express a strict preference for the intermediate choice *M*.²³

From the perspective of our theory this is not so surprising. Suppose that *S* is preferred by *u* and *L* by *V*. If *L* is chosen there is a high temptation so a high marginal

²³ We are grateful to Simone Cerreia-Vioglio for bringing the Camerer and Ho (1994) paper to our attention. They did not solicit willingness to pay for the most preferred option, so while stating indifference was an option, it may be that some subjects were indifferent but chose not to say so. Camerer and Ho (1994) conducted different sets of experiments with two populations: one of high school subjects and one of MBA students. The most relevant results are those with the MBA students, which is what we report here. Note that sixteen percent of subjects strictly prefer both *S* and *L* to the intermediate choice, which can be explained in our theory only with a decreasing marginal cost of self-control.

cost of self-control. With this high marginal cost of self-control, S may look quite attractive. If S then there is no temptation so a low marginal cost of self-control. This may make L attractive, suggesting that the optimum may lie in between.

To establish this formally, consider the utility of M as a function of λ :

$$\begin{aligned} U(\lambda) &= \lambda Eu(\tilde{c}_1^S) + (1 - \lambda)Eu(\tilde{c}_1^L) \\ &\quad - g(Eu(x_1 + \tilde{z}_1^S) - (\lambda Eu(\tilde{c}_1^S) + (1 - \lambda)Eu(\tilde{c}_1^L))) \\ &\quad + \delta[\lambda EV(w_2(\tilde{c}_1^S)) + (1 - \lambda)EV(w_2(\tilde{c}_1^L))]. \end{aligned}$$

Differentiating with respect to λ , we find

$$U'(\lambda) = (1 + \gamma_1(\lambda))(Eu(\tilde{c}_1^S) - Eu(\tilde{c}_1^L)) + \delta[EV(w_2(\tilde{c}_1^S)) - EV(w_2(\tilde{c}_1^L))],$$

where $\gamma_1(\lambda)$ denotes the first-period marginal cost of self-control as a function of λ . Notice that $\gamma_1(0) < \gamma_1(1)$ provided that g is strictly convex, that is, L has a higher marginal cost of self-control than S . This implies $\tilde{c}_1^S < \tilde{c}_1^L$ and because the problem is concave if $\gamma_1(0)$ is small enough $U'(0) > 0$, while if $\gamma_1(1)$ is large enough $U'(1) < 0$. These two conditions imply that there is some $0 < \lambda < 1$ such that M is strictly preferred to both S and L .

IV. Quantitative Analysis

We turn now to a quantitative analysis of the reversal phenomenon. Roughly speaking, the question we wish to address is whether there is a single set of preferences over risk, time, and self-control that can explain the behavior of each individual across a wide range of decisions. In doing so we need to contend with two issues: the heterogeneous nature of individuals, and the context in which decisions are made.

With respect to heterogeneity our focus here is on the median individual with the hope that this person does not differ too much from study to study. We do examine to a limited extent other people such as individuals who are more risk averse than the median.

With respect to the context in which decisions are made we make two adaptations to the model: we explicitly account for the fact that some consumer expenditure is precommitted (for durable goods and the like) and so not subject to temptation, and we drop the assumption that short-run and long-run risk aversion are linked.

Our methodology is that of calibration. We do not look at a particular collection of data on the Allais paradox and ask what parameter values might explain that data. Rather we use data from other sources to estimate behavioral parameters such as risk aversion and time preference, then ask whether these parameters are consistent with the Allais paradox. In this section we report the parameters we used; the sources of data and estimation procedures are reported in Appendix II.

Consumption Commitment.—So far, we have followed Fudenberg and Levine (2006). Now we consider an extension of that model that we will use to explain the

degree of risk aversion we observe in experimental data. Specifically, we suppose that in the bank sub-period there is a choice of nightclubs to go to in the nightclub sub-period. These choices are indexed by the quality of the nightclub $c_t^q \in (0, \infty)$. In a nightclub of quality c_t^q we assume that the utility of the short-run self has the form $u(c_t | c_t^q)$ depending on the amount consumed c_t there and the quality of the nightclub. We want this function to satisfy $u(c_t | c_t) = \log c_t$ so that in a deterministic and perfectly foreseen environment without self-control costs, behavior is the same as with standard logarithmic preferences. There are a great many possible functional forms satisfying these properties. Our choice of a specification is guided both by analytic convenience and by evidence (examined below) that short-term risk preferences seem more risk averse than consistent with the logarithmic specification, even when self-control costs are taken into account. This leads us to adopt the functional form²⁴

$$u(c_t | c_t^q) = \log c_t^q - \frac{(c_t/c_t^q)^{1-\rho} - 1}{\rho - 1},$$

where $\rho \geq 1$ corresponds to the short-run self's relative risk aversion over immediate consumption. Because our goal is a tightly parameterized model that fits the experimental data, this functional form assumes that all of the nightclubs have the same coefficient ρ . In practice ρ could vary. This might be important in accounting for the preferences of the very rich and the very poor; we discuss it in more detail when we examine the robustness of our calibrated parameters.²⁵

Durable Consumption.—The next step is to specify the agent's preferences for durable versus non-durable consumption. Durable consumption c_t^d can only be adjusted slowly, and seems unlikely to respond at all to the sorts of income shocks received in the lab experiments we study. For this reason, we simplify the model by assuming that the time path of c_t^d is chosen for once and for all in the initial time period.²⁶ Our goal here is simply to account for the fact that durable consumption exists, and not to explain it, so we adopt a simple Cobb-Douglas-like specification $\tau u(c_t | c_t^q) + (1 - \tau) \log c_t^d$; this will lead to a constant share $1 - \tau$ of spending on durables. From National Income and Product account data, we estimate $\tau = 0.57$. This number may be too large, as in addition to the durable categories we consider there are additional categories of consumption such as food, transportation, insurance, childcare, communication, and health club dues that also may not be subject to temptation. For that reason we will also consider a robustness check with a much smaller value of $\tau = 0.23$ calibrated so that $u(c_t | c_t^q)$ is the logarithm.

²⁴This functional form, while not implied by the basic assumptions of the dual-self approach, is a standard CES and is not chosen with our particular data in mind.

²⁵Note that $u(c_t | c_t^q) \leq u(c_t | c_t)$, so that when planning to consume a given amount c_t it is best to choose the nightclub of the same index. Intuitively, the quality c_t^q of a nightclub represents a "target" level of consumption expenditure at that nightclub, and people with different planned consumption levels will choose consumption sites with different characteristics. The quality of a nightclub can also be interpreted as a state variable or capital stock that reflects experience with a given level of consumption: a wine lover who unexpectedly wins a large windfall may take a while both to learn to appreciate differences in *grands crus* and to learn which ones are the best values.

²⁶Similarly, we abstract from labor supply, precautionary savings motives, and so forth. However we explicitly introduce durable consumption so that when we calibrate pocket cash the short-run self does not perceive that the rent check and similar expenses are available for short-term amusement.

Long-Run Self.—The objective function of the long-run self is

$$U_{RF} = E \sum_{t=1}^{\infty} \delta^{t-1} [\tau(u(c_t|c_t^q) - g(u(x_t|c_t^q) - u(c_t|c_t^q)) + (1 - \tau)\log c_t^d)],$$

which is to be maximized with respect to $c_t \geq 0$, $c_t^q \geq 0$, $c_t^d \geq 0$, $x_t \geq 0$ subject to w_{t+1} , given $w_{t+1} = R(s_t + x_t - c_t)$, $s_t + x_t + c_t^d \leq w_t$, and $w_t \geq 0$. Notice that this is a simple optimization problem with no uncertainty and perfect foresight. We provide solutions to this model in Appendix I. The key conclusion is that the consumption cutoff is given by

$$\hat{c}_1 \equiv x_1 + \hat{z}_1 = (x_1)^{(\rho-1)/\rho} \left[\frac{\tau(1 - \delta)}{\delta} (1 + \gamma_1)w_2 \right]^{1/\rho} \\ \approx x_1(1 + \gamma_1)^{1/\rho}.$$

Define $\mu_1(\gamma_1) = (1 + \gamma_1)^{1/\rho} \approx \hat{c}_1/x_1$. Because the numerical value of γ_1 is hard to interpret, in measuring the cost of self-control we will report $\mu_1(\gamma_1)$ rather than γ_1 .

Cost of Self-Control.—In our calibrations of the model, we assume that the cost function is quadratic: $g(v_t) = \gamma v_t + (1/2)\Gamma v_t^2$, and we will maintain that assumption throughout the rest of the paper. Thus, the marginal cost of self-control at time t is now $\gamma + \Gamma v_t$; we continue to denote this as γ_t .

Time Patience and Wealth.—We consider several plausible scenarios for the subjective discount factor, income, wealth, and pocket cash. For the time horizon Δ of the short-run self we use one day. For the purposes of robustness we checked that none of our results are sensitive to assuming a time horizon of a week: details can be found in the earlier working paper version available online.

Table 4 lists income corresponding to three different income levels,²⁷ and three different subjective interest rates, for which we take the base case to be 1 percent.

Risk Aversion.—In our theory as long as the combination of pocket cash and the maximum winning is well below our estimates of \hat{c}_1 , all income is spent, and our consumer simply behaves as a risk-averse individual with wealth equal to pocket cash.²⁸ In particular this means that the self-control parameters γ and Γ are irrelevant for the calibration, though they do help determine the magnitude of \hat{c}_1 . To determine the risk aversion of the short-run self, we examine small gambles using

²⁷ The middle level of income is the US median.

²⁸ Note that this theory predicts that if payoffs are delayed sufficiently, risk aversion will be much lower. Experiments reported in Nicholas Barberis, Ming Huang, and Richard H. Thaler (2006) suggest that there is appreciable risk aversion for gambles where the resolution of the uncertainty is delayed as well as the payoffs themselves. However, delayed gambles are subject to exactly the same self-control problem as regular ones, so this is consistent with our theory. In fact the number of subjects accepting the risky choice in the delayed gamble was in fact considerably higher than the non-delayed gamble, rising from 10 percent to 22 percent.

TABLE 4—CALIBRATED PARAMETER SUMMARY

Percent interest r		$y_1 = 14\text{K}$		$y_1 = 28\text{K}$		$y_1 = 56\text{K}$	
Annual	Daily	w_1	x_1	w_1	x_1	w_1	x_1
1	0.003	1.3M		2.6M		5.2M	
3	0.008	0.43M	20	0.86M	40	1.7M	80
5	0.014	0.30M		0.61M		1.2M	

TABLE 5—ESTIMATED RELATIVE RISK AVERSION

	Pocket cash x_1		
	\$20	\$40	\$80
ρ median	1.17	1.43	1.90
ρ 85 th	2.02	2.74	4.15

data from Holt and Laury (2002). In Appendix II we use this data to estimate the coefficient of relative risk aversion ρ that explains the behavior of the median individual and the coefficient that matches the behavior of the individual at the 85th percentile of risk aversion.

Our estimates of median and “high” risk aversion are given in Table 5.

V. Quantitative Analysis of Reversals

We proceed next to examine the Allais paradox and other reversals in the calibrated model. We assume that the choice in these experiments is completely unanticipated. In this case the solution is simple: there is no self-control problem at the bank, so the choice is $c_1^* = x_1$ and spend all the pocket cash in the nightclub of choice. Given this, the problem is purely logarithmic, so the solution is to choose $x_1 = (1 - \delta)w_1$.

In the Kahneman and Tversky (1979) version of the Allais paradox, the two options in the first scenario are L_1 given by (0.01:0, 0.66:2,400, 0.33:2500), while S_1 is 2,400 for certain. Many people choose option S_1 . In scenario two the pair of choices are $L_2 = (0.67:0, 0.33:2,500)$ and $S_2 = (0.66:0, 0.34:2,400)$.²⁹ Here many people choose L_2 . Expected utility theory requires the same option L or S be chosen in both scenarios.

The original Allais paradox involved substantially higher stakes, so it would be difficult to implement other than as a thought experiment: option L_1 was (0.01:0, 0.89:1,000,000, 0.1:5,000,000) and S_1 was 1,000,000 for certain; the second scenario was $L_2 = (0.90:0, 0.10:5,000,000)$ and $S_2 = (0.89:0, 0.11:1,000,000)$.

From our qualitative analysis, we know that to get a reversal the safe alternative S must be preferred by u and the risky alternative L must be preferred by V . The

²⁹ These were thought experiments. We are unaware of data from Allais experiments with real payments of over \$2,000, though other experiments with stakes of analogous shares of average per-capita yearly income have been conducted in poor countries. There is experimental data on the Allais paradox with real, but much smaller, stakes; we discuss this below. We thank Chew Soo Hong for drawing our attention to an error in our discussion of this version of the paradox in a previous draft.

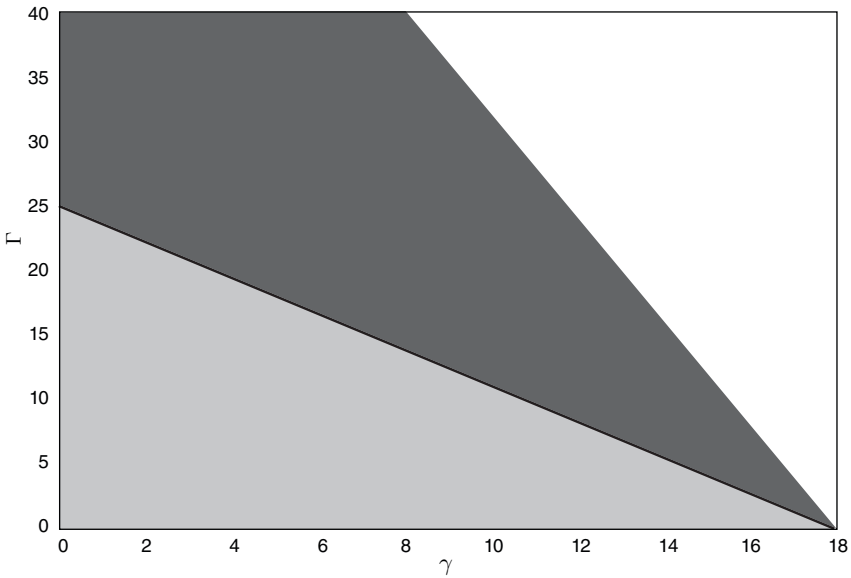


FIGURE 2. ALLAIS PARADOXES

first question is whether this is true for the calibrated parameters. A computation shows that it is. However, the data cannot easily pin down the curvature of the self-control function, beyond the fact that it cannot be linear. This is shown in Figure 2, where the (γ, Γ) pairs generating Allais paradoxes are computed for the base case of \$28,000 annual income and risk aversion $\rho = 1.43$. In the light-shaded region with low costs of self-control, the long-run optimum L is the best choice in both scenarios. In the unshaded region with high costs of self-control, the short-run optimum S is the best choice in both scenarios. In the dark-shaded region in between an Allais reversal occurs, as the optimal choice is S in the high temptation scenario and L in the low temptation scenario.

While a wide range of curvatures of the self-control function is consistent with the data, the marginal cost of self-control must lead to near indifference and so is well pinned down by the data. In Table 6 we report the types of marginal costs that are consistent with the Allais paradox and other reversals, using the median and 85 percent risk aversion estimates and typical income values from the last section.³⁰

Note that the estimates of the threshold μ_1 depend on the preference parameters such as risk aversion and pocket cash. This is a necessary feature of the theory: since the data to be explained is the same, if we change other preference parameters, the self-control cost must also change to be consistent.

The crucial quantitative issue is the reasonableness of the parameters and their consistency with other data and other types of experiments. From the table, we can see that the data from the Allais experiments are consistent with the level of pocket cash and risk aversion in the Rabin paradox. This is a non-trivial quantitative finding,

³⁰ The programs used for the computations were in Octave, a free equivalent of Matlab. They can be found at www.dklevine.com/papers/allais.zip.

TABLE 6—SELF CONTROL ESTIMATES (KT = *Kahneman and Tversky Allais*; OA = *Original Allais*;
CR = *Common Ratio*^a; CL = *Cognitive Load*^b)

Case	Stakes	r	τ	Income	$x_1 = c_1^*$	ρ	μ_1
KT	2,400	1 percent	0.57	14,000	20	1.17	16.7
KT	2,400	1 percent	0.57	14,000	20	2.02	7.56
KT	2,400	1 percent	0.57	28,000	40	1.43	8.49
KT	2,400	1 percent	0.57	28,000	40	2.74	4.13
KT	2,400	1 percent	0.57	56,000	80	1.90	4.57
KT	2,400	1 percent	0.57	56,000	80	4.15	2.47
KT	2,400	1 percent	0.23	28,000	16.4	1.00	24.3
OA	5,000,000	1 percent	0.57	28,000	40	1.43	4,340
OA	5,000,000	1 percent	0.57	28,000	40	2.74	128
KT	2,400	3 percent	0.57	14,000	20	2.02	7.34
CR	12	3 percent	0.57	14,000	20	2.02	1.89
CL	1.0	7 percent	1.0	11.4	2.3	1.17	21.2

Notes:

^aData from Baucells and Heukamp (2010a) as described earlier.

^bData from Benjamin, Brown, and Shapiro (2006); the experiment and data are described in the Appendix.

as to generate a reversal the long-run and short-run self must have opposite preferences over the alternatives, and the two selves do not have opposite preferences over these choices for all combinations of pocket cash and risk aversion. Moreover, the preferences of the long-run and short-run self must be sufficiently close to indifference that we find marginal cost of self-control parameters that generate exact indifference. Again, this is true for all the reversals examined.

In terms of the reasonableness of the calibrated parameters, the key issue is the range of marginal self-control costs μ_1 that are needed to generate reversal. The high stakes Allais paradoxes require very large marginal costs of self-control: in both of these “OA” cases, this is because the cutoff at which self-control is used must lie below but close to the amount that is won. This is qualitatively consistent with the theory, in the sense that with convex cost of self-control higher stakes do lead to higher marginal cost of self-control, but it is not possible to fit the data with a single quadratic marginal cost of self-control function. It should be noted that the high stakes here are very high, and it may not be reasonable to approximate long-run preferences over such a large range with a logarithmic utility function.

More broadly, the parameters here have difficulty with other data that involves much higher or much lower stakes. They are hard to reconcile with the very low stakes in the common consequence case, and they cannot explain the Allais paradox for very small stakes: When all payoffs are below pocket cash the decision-maker simply uses the preferences of the short-run self, and so behaves as an expected utility maximizer with wealth equal to pocket cash. However, whether the Allais paradox occurs for very small stakes is controversial.³¹

³¹ Raymond Battalio, John H. Kagel, and Komain Jiranyakul (1990) found that subjects did exhibit the Allais paradox even for very small stakes in the range \$0.12 to \$18.00, and that they less frequently even made the reverse Allais choices. However, indifference or near indifference may be a key factor in the reported results. Glenn W. Harrison (1994) found that with various small stakes the Allais paradox was sensitive to using real rather than hypothetical payoffs, and found in the real payoff case only 15 percent of the population exhibited the paradox, and a

Finally, there is evidence on the level of self-control costs for much larger stakes from survey data on the response of consumption to windfall income such as inheritances and stimulus tax rebates. Most of the literature we have examined finds a high propensity to consume out of small or modest windfalls, but does not examine whether this propensity drops for larger windfalls, so it is difficult to pin down the size of a possible income threshold. For example, Jonathan A. Parker et al. (2010) find a substantial marginal propensity to consume from 2008 economic stimulus payments that ranged from \$300–\$600 for single individuals and \$600–\$1,200 per couple, with additional payments of \$300 per “qualified” child. The large marginal propensity to consume from these payments suggests that the threshold for self-control is larger than say 5 percent, corresponding to μ_1 of about 35, but does not pin down the payment level at which self-control would be used to lower the fraction of the payment consumed.³² The one study we know of that uses income measures to estimate this cutoff is Mohamed Abdel-Ghany et al. (1983); they find a cutoff of 10 percent of income which corresponds to μ_1 of about 70. Below this threshold the marginal propensity to consume out of income is 94 percent; above this threshold it is only 2 percent. This is a much higher value of μ_1 than we find in all but the high stakes Allais cases.

VI. Conclusion

Because the long-run self in the dual-self model faces a dynamic programming problem, the model can be applied to a wide range of problems that combine choices over time with uncertain outcomes. Moreover, we have argued that a simple version of the model with a quadratic cost of self-control and logarithmic preferences for foreseen consumption is not only flexible enough to allow for various sorts of paradoxical behavior but can also quantitatively fit a range of data.

One class of alternative models that can potentially be used to explain a wide range of phenomena are probability weighting models, including such prospect theory and models of ambiguity aversion. These models are typically static, and while they often have a reference point that can be used, for example, to resolve the Rabin paradox, that reference point is generally exogenously given.³³ However, there are recent efforts to extend these models to examine dynamic phenomena. For example, Halevy (2008) and Baucells and Heukamps (2010b) propose models of choice over time paths of money that permit the attenuation of the Allais paradox by delay and the attenuation of immediacy by uncertainty. However, the only probability weighting model that has been studied extensively is that of prospect theory.³⁴

following study by Michael S. Burke et al. (1996) found a statistically significant drop from 36 percent to 8 percent. John Conlisk (1989) also finds little evidence of an Allais paradox when the stakes are small.

³² A similar issue arises in interpreting the findings of David S. Johnson, Parker, and Nicholas S. Souleles (2006) and a number of related studies.

³³ See Botond Köszegi and Rabin (2006) for one way to make the reference point endogenous, and Gul and Pesendorfer (2008) for a critique.

³⁴ There are many other alternative explanations of specific paradoxes, for example regret theory, but these are less comprehensive than either this theory or prospect theory. There are also paradoxes that are hard to explain in any theory. For example Justin Sydnor (2009) shows that a typical homeowner pays \$100 to reduce his deductible from \$1,000 to \$500 with a claim rate of less than 5 percent. Rather than revealing something about risk preference, this may be the result of the homeowner's not knowing the probability of a loss.

Despite the fact that prospect theory has been used for many years, few efforts have been made to examine whether individual behavior over a range of different paradoxes can be explained in a single consistent way. William Neilson and Jill Stowe (2002) conducted a systematic examination of the parameters needed to fit cumulative prospect theory (with rank-dependent probabilities) to various empirical facts, and concluded that “there are no parameter combinations that allow for both the desired gambling/insurance behavior and a series of choices made by a strong majority of subjects and reasonable risk premia.”³⁵

The point is simply that prospect theory does not explain the data used to motivate it with a single consistent set of preferences. Self-control theory, as examined, faces a similar difficulty, although arguably less severe.

Self-control theory can address issues such as the role of cognitive load and explains intertemporal paradoxes such as the hyperbolic discounting phenomenon about which prospect theory is generally silent. Also, our approach provides a self-contained theory of intertemporal decision making; by way of contrast, while it is not transparent how to embed prospect theory into an intertemporal model.^{36,37}

In the other direction, prospect theory allows for individuals who are simultaneously risk averse in the gain domain and risk loving over losses. This is done in part through the use of different value functions in the gain and loss domains, and in part through its use of a probability weighting function, which can allow individuals to overweight rare events.³⁸ Most work on prospect theory has estimated a representative-agent model; Adrian Bruhin, Helga Fehr-Duda, and Thomas Epper (2010) refined this approach by classifying individuals as expected utility maximizing or as cumulative prospect theory types,³⁹ and find that most individuals are prospect theory types.

There are two main deficiencies we have identified in current versions of the dual-self model. First, there is a tension with respect to the degree of self-control. The model predicts that there should be a threshold level of unanticipated income, with marginal propensity to consume of 100 percent below the threshold and a very low marginal propensity above it. However, it is difficult to reconcile the range of cut-offs needed to explain the data with the behavior of single individual. Second, as is the case with quasi-hyperbolic discounting, the model predicts very abrupt

³⁵ The paper also says that “the preference function estimated by Tversky and Kahneman (1992) can accommodate neither the strongest choice patterns from Battalio, Kagel, and Jiranyakul (1990) nor the Allais paradox, and implies some rather large risk premia. The preference functions estimated by Camerer and Ho (1994) and George Wu and Richard Gonzalez (1996) imply virtually no risk seeking over unlikely gains and virtually no risk aversion over unlikely losses, so that individuals will purchase neither lottery tickets nor insurance.”

³⁶ Köszegi and Rabin (2009) develop but do not calibrate a dynamic model of reference dependent choice.

³⁷ We have also argued that the same model can account for risky decision making of Chilean high school students faced with differing cognitive loads. This finding is not trivial, as there are possible observations that are not consistent with the theory. For example, cognitive load in the Chilean experiment could have caused subjects to switch in the reverse, “anti-Allais,” direction, which our model could not explain. Second, there is enormous heterogeneity in the data; only a fraction of subject populations exhibit reversals, and the populations in the various experiments are very different.

³⁸ See Drazen Prelec (1998) for an axiomatic characterization of several probability weighting functions, and a discussion of their properties and implications.

³⁹ Their estimation procedure tests for and rejects the presence of additional types. Given the functional forms they estimate, individuals with expected utility preferences are assumed to be risk averse throughout the gains domain, while in their data individuals are risk loving for small probabilities of winning, while for higher probability of success they are risk averse.

responses to small changes in the timing of rewards near the end of the time corresponding to the end of the period, yet this is inconsistent both with common sense and a great deal of data.

Nevertheless, despite the fact that it has so far received less attention and development, the dual-self model explains choices over lotteries about as well as the considerably more developed prospect theory, while explaining phenomena such as commitment and cognitive load that prospect theory cannot. Moreover, the dual-self model is a fully dynamic model of intertemporal choice that is consistent with both traditional models of savings (long-run logarithmic preferences) and with the equity premium puzzle.⁴⁰ As these models of individual behavior are widely and successfully used by macroeconomists (including New Keynesians) to explain a wide variety of aggregate data, ranging from bubbles to exchange rate paradoxes and the equity premium puzzle, it is important not to simply discard all this existing work.

Further improvements in the dual-self model are desirable, but its success in providing a unified explanation for a wide range of phenomena suggests that it should be viewed as a natural starting point for attempts to explain other sorts of departures from the predictions of the standard model of consumer choice. We hope here to convince the economics community of the importance of quantitative calibrations of decision theory models. One possible next step would be try to more explicitly account for the evident heterogeneity of the population, and estimate distributions of self-control parameters as opposed to simply fitting the median or some other fractile as we have done here. Another direction is to further extend the theory, as we do in Fudenberg and Levine (2010) by extending the model to account for the fact that the effect of delay is gradual as opposed to the abrupt effect in both the dual-self model and quasi-hyperbolic utility. This more complicated model should also be calibrated to experimental data, but this research has yet to be done.

APPENDIX I: THEORY

The text considers four cases: the *general case* of nested utility and durable consumption, and two special cases: *nested CES/log*, the *base case* of non-nested utility without durable consumption, and the *base case specialized to logarithmic* utility. Here we prove the results stated in the text, with some generalizations.

⁴⁰ The “behavioral life cycle model” of Hersh M. Shefrin and Thaler (1988) can also explain many qualitative features of observed savings behavior, and pocket cash in our model plays a role similar to that of “mental accounts” in theirs. The behavioral life cycle model takes the accounts as completely exogenous, and does not provide an explanation for preferences over lotteries. It does seem plausible to us that some forms of mental accounting do occur as a way of simplifying choice problems. In our view this ought to be derived from a model that combines the long-run/short-run foundations of the dual-self model with a model of short-run player cognition.

Calculation of $V_1(w_1)$ and $V_2(w_2|w_1)$ with Nested CES/Log.—Recall from the text that $x_t = (1 - \delta)\tau w_t$, $c_t^d = (1 - \tau)(1 - \delta)w_t$, $w_{t+1} = R(w_t - x_t - c_t^d)$ so $w_{t+1} = R\delta w_t$ and thus $w_t = (R\delta)^{t-1}w_1$. We may then compute

$$\begin{aligned} V_1(w_1) &= \sum_{t=1}^{\infty} \delta^{t-1} [\tau \log x_t + (1 - \tau) \log c_t^d] \\ &= \sum_{t=1}^{\infty} \delta^{t-1} [\tau \log((1 - \delta)\tau w_t) + (1 - \tau) \log((1 - \tau)(1 - \delta)w_t)] \\ &= \sum_{t=1}^{\infty} \delta^{t-1} [\log(1 - \delta) + \log((R\delta)^{t-1}w_1) + \tau \log(\tau) \\ &\quad + (1 - \tau) \log(1 - \tau)] \\ &= \frac{\log(w_1)}{(1 - \delta)} + \frac{1}{(1 - \delta)} \left(\log(1 - \delta) + \tau \log(\tau) \right. \\ &\quad \left. + (1 - \tau) \log(1 - \tau) + \frac{\delta \log(R\delta)}{(1 - \delta)} \right). \end{aligned}$$

Turning to $V_2(w_2|w_1)$, we observe that that c_t^d is already committed in the first period, the present value of utility from durable consumption starting in period 2 is a constant, it follows that the agent's problem reduces to maximizing

$$\tau \sum_{t=2}^{\infty} \delta^{t-2} \log x_t$$

subject to $w'_2 \equiv w_2 - (1 - \tau)R\delta w_1$ and $w'_{t+1} = R(w'_t - x_t)$. The solution is $x_t = (1 - \delta) \times w'_t$, so the general form of the maximized present value is $(\tau \log(w'_t)/(1 - \delta)) + K''$, where K'' is a constant independent of w_1 . Substituting $w'_2 \equiv w_2 - (1 - \tau)R\delta w_1$ we see that the maximized present value from non-durable consumption is

$$\frac{\tau \log(w_2 - (1 - \tau)R\delta w_1)}{1 - \delta} + K''.$$

The present value of the utility from durable consumption is

$$\frac{(1 - \tau) \log((1 - \tau)w_1)}{1 - \delta} + K'''$$

where K''' , like the other constants, is independent of w_1 . Thus the overall present value of utility beginning in period 2 is

$$\begin{aligned} V_2(w_2|w_1) &= \frac{\tau \log(w_2 - (1 - \tau)R\delta w_1)}{1 - \delta} \\ &\quad + \frac{(1 - \tau) \log((1 - \tau)R\delta w_1)}{1 - \delta} + K'. \end{aligned}$$

Using the fact that $V_2(\delta R w_1 | w_1) = V_1(\delta R w_1)$, we see that

$$\begin{aligned}
 V_2(R\delta w_1 | w_1) &= \frac{\tau \log(R\delta\omega_1 - (1 - \tau) R\delta w_1)}{1 - \delta} \\
 &\quad + \frac{(1 - \tau) \log((1 - \tau) R\delta w_1)}{1 - \delta} + K' \\
 &= \frac{\tau \log(\tau) + (1 - \tau) \log(1 - \tau) + \log(R\delta\omega_1)}{1 - \delta} + K' \\
 &= \frac{\log(R\delta w_1)}{(1 - \delta)} + \frac{1}{(1 - \delta)} \left(\log(1 - \delta) + \tau \log(\tau) \right. \\
 &\quad \left. + (1 - \tau) \log(1 - \tau) + \frac{\delta \log(R\delta)}{(1 - \delta)} \right),
 \end{aligned}$$

so that

$$K' = \frac{1}{1 - \delta} \left[\log(1 - \delta) + \frac{\delta \log(R\delta)}{(1 - \delta)} \right].$$

We prove Theorem 1 in the general case. The text makes use of the nested CES/log and of the base case.

THEOREM 1: *For given (x_1, c_1^*) and each $j \in \{S, L\}$ there is a unique solution to*

$$\gamma_1^j = \hat{\gamma}_1^j(\gamma_1^j).$$

This solution together with $\tilde{c}_1^j = \min(\hat{c}_1^j(\gamma_1^j)(z_1^j), x_1 + z_1^j)$ and the choice of j that maximizes long-run utility is necessary and sufficient for an optimal solution.

PROOF:

Consider random unanticipated income \tilde{z}_1^j at the nightclub. If z_1 is the realized income, the short-run self is constrained to consume $c_1 \leq x_1 + z_1$. Period 2 wealth is given by

$$w_2 = R(s_1 + x_1 + z_1 - c_1 - c_1^d) = R(w_1 + z_1 - c_1 - c_1^d).$$

The utility of the long-run self starting in period 2 is given by the solution of the problem without self-control.

Let \tilde{c}_1 be the optimal response to the unanticipated income \tilde{z}_1 . This is a random variable measurable with respect to \tilde{z}_1 . The overall objective of the long-run self is to maximize

$$(A1) \quad \tau(Eu(\tilde{c}_1^j | c_1^q) - \bar{g}(x_1, \tilde{c}_1^j, c_1^q)) \\ + \frac{\delta}{(1 - \delta)} EV(R(w_1 + \tilde{z}_1^j - \tilde{c}_1^j - c_1^d) | w_1),$$

where \bar{g} is the control cost associated with the specified pocket cash and non-durable consumption. Let $\bar{u}_1(x_1 | c_1^q) = \max\{Eu(x_1 + \tilde{z}_1^S | c_1^q), Eu(x_1 + \tilde{z}_1^L | c_1^q)\}$ denote the maximum possible utility given c_1^q and the pair of lotteries S, L . Since \bar{u}_1 does not depend on \tilde{c}_1^j , the optimal level of consumption can be determined for each lottery realization by pointwise maximization of (A1) with respect to $c_1 = c_1^j(z_1^j)$. The value of the marginal cost of self-control is given by

$$(A2) \quad \gamma_1 = g'(\bar{u}_1(x_1 | c_1^q) - Eu(\tilde{c}_1^j | c_1^q)) \\ = g'(\bar{u}_1(x_1 | c_1^q) - \sum_{z_1^j} \Pr(z_1^j) u(c_1^j(z_1^j) | c_1^q)).$$

First, we show that the first-order conditions corresponding to optimal consumption for a given choice j have a unique solution. Observe that

$$\frac{d\gamma_1}{dc_1^j(z_1^j)} = -\Pr(z_1^j) \frac{\partial u(c_1^j(z_1^j) | c_1^q)}{\partial c_1^j} g''(\bar{u}_1(x_1 | c_1^q) - Eu(\tilde{c}_1^j | c_1^q)) \leq 0.$$

The derivative of (A1) with respect to $c_1^j = c_1^j(z_1^j)$ evaluated at z_1^j is

$$\tau(1 + \gamma_1) \frac{\partial u(c_1^j | c_1^q)}{\partial c_1^j} - \frac{\delta}{(1 - \delta)} \frac{\partial V(R(w_1 + z_1^j - c_1^j - c_1^d) | w_1)}{\partial c_1^j}.$$

From this we can compute the second derivative

$$\tau \frac{\partial u(c_1^j | c_1^q)}{\partial c_1^j} \frac{d\gamma_1}{dc_1^j} + \tau(1 + \gamma_1) \frac{\partial^2 u(c_1^j | c_1^q)}{\partial c_1^j{}^2} \\ + \frac{\delta}{(1 - \delta)} \frac{\partial^2 V(R(w_1 + z_1^j - c_1^j - c_1^d) | w_1)}{\partial c_1^j{}^2} < 0,$$

where the first term is non-positive and each other term is strictly negative, implying that the function is globally concave.

Because the objective function is globally concave with respect to c_1^j , it follows that the unique maximum is given by the solution to the first-order condition.

We now show that the conditions in Theorem 1 are necessary and sufficient for an optimum. Since the problem is one of maximizing a continuous function over a compact space, an optimum exists. To show necessity, consider any optimum, and suppose the choice is j . Then for any given consumption plan in j the marginal cost of self-control γ_1 is defined by (A2), and the optimal consumption plan must satisfy the first-order condition with respect to that γ_1 because our conditions preclude a boundary solution. That is, $\gamma_1^j = \hat{\gamma}_1^j(\gamma_1^j)$ must hold.

Next we show sufficiency. Suppose j, γ_1^j satisfy the conditions of the theorem and that this is not the optimum. That optimum must yield more utility than choosing $-j$ and any consumption plan in $-j$, so the unique consumption plan that comes from solving $\gamma_1^{-j} = \hat{\gamma}_1^{-j}(\gamma_1^{-j})$. Given that j is chosen, the optimal consumption is the unique solution of the first-order condition. On the other hand, if $-j$ was chosen, we could do no better than the consumption plan defined by $\gamma_1^{-j} = \hat{\gamma}_1^{-j}(\gamma_1^{-j})$, and by assumption this is not as good as choosing j .

Computations in the Nested CES/Logarithmic Case.—The first-order conditions can be written as

$$\begin{aligned} (c_1^j)^\rho &= (c_1^q)^{\rho-1} \frac{\tau(1-\delta)(1+\gamma_1)}{\delta} (w_1 + z_1^j - c_1^j - c_1^d) \\ &= \bar{K}(w_1 + z_1^j - c_1^j - c_1^d). \end{aligned}$$

Since the left-hand side of this equality is continuous and increasing in c_1^j , and is 0 when c_1^j is 0, while the right-hand side is linearly decreasing in c_1^j starting from a positive value, this equality has exactly one solution; moreover the solution is increasing in \bar{K} , (and so decreasing in δ and increasing in γ_1), and increasing in $w_1 + z_1^j$. We prove Theorem 2 for the base case used in the text.

THEOREM 2: *Let $\xi \equiv \sup_{c_1}(-c_1 u''(c_1)/u'(c_1))/\inf_{c_1}(-c_1 u''(c_1)/u'(c_1))$. Then*

$$-\frac{V''(w_1)}{V'(w_1)} \leq -\xi \frac{c_1}{w_1} \frac{u''(c_1)}{u'(c_1)}.$$

PROOF:

Let c_t be the optimal plan when initial wealth is w_1 . For $z \geq 0$ define

$$U(z) \equiv \sum_{t=1}^{\infty} \delta^{t-1} u(c_t + (c_t/w_1)z).$$

Notice that the budget constraint implies that

$$\sum_{t=1}^{\infty} R^{-t+1} (c_t/w_1) = 1,$$

so that if z is an increment to wealth the plan $c_t + (c_t/w_1)z$ is feasible and satisfies the budget constraint with equality. This implies first that $V(w_1 + z) \geq U(z)$, and second, from the first-order conditions that $V'(w_1) = U'(0)$.

Since V is concave, these two facts imply that V is twice differentiable at w_1 and that $-V''(w_1) \leq -U''(0)$.

To finish the proof, write out

$$\begin{aligned} -U''(0) &= -\sum_{t=1}^{\infty} \delta^{t-1} u''(c_t)(c_t/w_1)^2 \\ &= (1/w_1) \sum_{t=1}^{\infty} \delta^{t-1} u''(c_t)(c_t/w_1) \{-c_t u''(c_t)/u'(c_t)\} \\ &\leq (\sup_{c_t} \{-c_t u''(c_t)/u'(c_t)\}/w_1) \sum_{t=1}^{\infty} \delta^{t-1} u''(c_t)(c_t/w_1) \\ &= (\sup_{c_t} \{-c_t u''(c_t)/u'(c_t)\}/w_1) U'(0). \end{aligned}$$

Since by the first-order conditions of dynamic programming $V'(w_1) = u'(c_1)$, we may write

$$-V''(0) \leq (\sup_{c_t} \{-c_t u''(c_t)/u'(c_t)\}/w_1) V'(0)$$

and the result follows by noting that

$$\frac{-c_1 u''(c_1)/u'(c_1)}{\inf_{c_t} \{-c_t u''(c_t)/u'(c_t)\}} \geq 1.$$

APPENDIX II: RISK AND THE RABIN PARADOX

We use data from Holt and Laury (2002), who did a careful laboratory study of risk aversion. Their subjects were given a list of ten choices between two lotteries S and L . The specific lotteries are shown below, where the first four columns show the probabilities of the rewards, and the first four rows, which are irrelevant to our analysis, are omitted.

Initially subjects were told that one of the ten rows would be picked at random and they would be paid the amount shown. Then they were given the option of renouncing their payment and participating in a high stakes lottery, for either 20X, 50X, or 90X of the original stakes, depending on the treatment. The high-stakes lottery was otherwise the same as the original: a choice was made for each of the ten rows, and one picked at random for the actual payment. Everyone in fact renounced their winnings from the first round to participate in the second. The choices made by subjects are shown in Table A1. Here we have marked (by asterisk and double-asterisk

TABLE A1—LABORATORY PREFERENCES TOWARDS RISK

Option <i>S</i>		Option <i>L</i>		Fraction of subjects choosing <i>S</i> ^a			
\$2.00	\$1.60	\$3.85	\$0.10	1X	20X	50X	90X
Number of observations =>				(187)	(150)	(19)	(18)
0.5	0.5	0.5	0.5	0.70(0.03) 4.2	0.85(0.03)	1.0(0.0)	0.90(0.07)
0.6	0.4	0.6	0.4	0.45(0.04) 12*	0.65(0.04) 1.1	0.85(0.08)	0.85(0.09)
0.7	0.3	0.7	0.3	0.20(0.03) 21**	0.40(0.04) 1.9*	0.60(0.12) 1.2	0.65(0.12)
0.8	0.2	0.8	0.2	0.05(0.02) 32	0.20(0.03) 2.8**	0.25(0.1) 1.8*	0.45(0.12) 1.5
0.9	0.1	0.9	0.1	0.02(0.01) 50	0.05(0.02) 4.2	0.15(0.08) 2.6**	0.40(0.12) 2.1*
1.0	0.0	1.0	0.0	0.00(0.0)	0.00(0.0)	0.00(0.0)	0.00(0.0)

Note: Bold numbers are coefficients of relative risk aversion giving indifference when $x_1 = 40$.

^aStandard errors in parentheses.

respectively⁴¹) the cells that correspond most closely to the median and eighty-fifth percentiles.

Assuming that the amounts of the lotteries lie below the threshold \hat{z}_1 , for each choice S, L and for given pocket cash x_1 we can compute the coefficient of relative risk aversion ρ that leads to indifference between the two choices.

$$E(x_1 + \tilde{z}_1^S)^{1-\rho} = E(x_1 + \tilde{z}_1^L)^{1-\rho}.$$

These numbers are reported in bold face also in Table A1 for the base level of pocket cash $x_1 = 40$.

We use Table A1 to infer the sample median and the 85th percentile of risk aversion. For example, we can infer from the 1X column that the median individual prefers S when the probability of the high payment is 0.5, but prefers L when the probability of the high payment is 0.6. Hence the sample median value of ρ must lie between 4.2, which would lead to indifference when the probability is 0.5, and 12, which would lead to indifference when the probability is 0.6. Table A2 gives the ranges for the median value of risk aversion as well as the 85th percentile based on the data in Table A1.

The first thing to observe is that the ranges for the 20X, 50X, and 90X treatments are generally consistent with each other. The 1X treatments exhibit considerably higher risk aversion. This cannot be due to sampling error: Taking a relatively broad confidence band of three standard deviations, we see that 70 percent of subjects choose S in the 50-50 1X cell, three standard deviations below the mean is still 61 percent of subjects, so that the median relative risk aversion should be at least the 4.2 reported for that cell. On the other hand, in the 20X treatment, we see 20 percent of subjects choos-

⁴¹ Asterisk for the median; double-asterisk cells for the eighty-fifth percentile.

TABLE A2—BOUNDS FOR QUANTILES OF COEFFICIENT OF RELATIVE RISK AVERSION

Stakes	Pocket cash $x_1 = 40$			
	50th percentile risk aversion		85th percentile risk aversion	
1X	4.2	12	21	32
20X	1.1	1.9	2.8	4.2
50X	1.2*	1.8	2.6	2.6
90X		1.5**	2.1	

ing S in the 80-20 20X cell, and here three standard deviations above the mean is only 29 percent of subjects, so that the median should be no more than the 2.8 reported for that cell, well below the 4.2 minimum estimate for the 1X treatment. We conclude that the CES functional form does not fit very small stakes gambles particularly well. This is well known from the empirical prospect theory literature, but is not a major concern, as we are not going to study gambles in the lower 1X range.⁴²

Examining the 20X, 50X, and 90X ranges, for the median, we see that we can pin down the coefficient of relative risk aversion in the sample between a low value of 1.2 (single asterisk) and a high value of 1.5 (double asterisk). Taking a non-parametric approach to estimating the corresponding population parameter in this case is unproblematic since the sample median is a consistent estimator of the population median.⁴³

For the 85th percentile—people who are more risk averse than 85 percent of the population—the sample data is contradictory, as in the 50X treatment the coefficient of relative risk aversion is pinned down to 2.6, while the range in the 20X treatment is 2.8 to 4.2. However, the 50X treatment is subject to substantial sampling error: the estimate the 15 percent of subjects choosing S in 90-10 50X cell has a standard error of 0.08, so the information in that cell has little value. Hence the non-parametric approach gives a range from 2.8 to 4.2 for the 85th percentile.

To choose a specific value from the ranges in Table A1, we choose values that fit these single asterisk (double asterisk) cells as closely as possible. We use the CES utility scaled by the pocket cash constant x_1

$$-x_1 \frac{(c_1/x_1)^{1-\rho} - 1}{\rho - 1}.$$

For any given coefficient of relative risk aversion and any cell in Table 4, we can compute a squared utility difference between S and L . For the bold face coefficients shown in Table 4, this number will be zero, but we cannot choose a single value to simultaneously make the utility difference in all the cells equal to zero. Instead we compute the value of ρ that minimizes the sum of these squared utility differences. This yields a median of 1.43, which lies in the range 1.2 to 1.5 that we have already

⁴² It is possible that the size of the choices might have been confounded with the order in which the choices were given. Harrison et al. (2005) find that corrected for order the impact of the size of the gamble is somewhat less than Holt and Laury found, a point which Holt and Laury (2005) concede is correct. The follow-on studies which focus on the order effects do not contain sufficient data for us to get the risk aversion estimates we need.

⁴³ We could use a parametric model by assuming an explicit functional form for the population distribution of risk aversion, but such an estimator would be susceptible to specification error.

identified, and an 85th percentile of 2.74, which lies slightly below the range from 2.8 to 4.2.⁴⁴ We carry out the same computation for our other candidates for pocket cash \$20 and \$80, with the results reported in Table 5. Because the estimated ρ s are greater than 1, these preferences are not logarithmic. Notice that the data does not let us separately identify pocket cash and risk aversion; various combinations of these two are observationally equivalent.

We also wish to consider values of τ smaller than 0.57.

Since we do not allow less risk aversion than the logarithm, the smallest possible value of pocket cash is that which gives rise to $\rho = 1$. We calibrate this to be $x = 16.4$, which corresponds to a value of $\tau = 0.23$ with the median income level. We use this value in robustness checks below.

APPENDIX III: COGNITIVE LOAD

To model cognitive load we assume that the control cost g depends on the sum of the foregone utility and the cognitive load; formally, if the load in period t is denoted d_t then the cost of self-control in period t is $g(\bar{u}_t - u_t + d_t)$. The theory predicts that increasing cognitive load should increase the marginal cost of self-control and lead to reversals similar to those in the Allais paradox. Relatively few experiments have been conducted on the effect of cognitive load on decisions involving risk. One recent one is an experiment conducted with Chilean high school juniors by Benjamin, Brown, and Shapiro (2006). This shows Allais type reversals generated by changes in cognitive load rather than changes in scenarios.⁴⁵

In the experiment students made choices about uncertain outcomes both under normal circumstances and under the cognitive load of having to remember a seven-digit number while responding. In scenario 1 the choice was between a 50-50 gamble between 650 pesos and nothing versus a sure option of 250 pesos. In scenario 2 the sure option was replaced by a 50-50 gamble between 300 and 200 pesos.⁴⁶ Table A3 summarizes the fraction of the population taking the risky choice, with the number in parentheses following the treatment indicating the number of subjects. These were real, and not hypothetical choices, the subjects were paid in cash at the end of the session. To provide some reference for these

⁴⁴ But also in the direction of the 90-10 50X cell.

⁴⁵ The main focus of Benjamin, Brown, and Shapiro (2006), like that of Frederick (2005), is on the correlation between measures of cognitive ability and the phenomena of small-stakes risk aversion and of a preference for immediate rewards. Benjamin, Brown, and Shapiro (2006) find a significant and substantial correlation between each of these sorts of preferences and cognitive ability. They also note that the correlation between cognitive ability and time preference vanishes when neither choice results in an immediate payoff, and that the correlation between small-stakes risk aversion and "present bias" drops to zero once they control for cognitive ability. This evidence is consistent with our explanation of the Rabin paradox, as it suggests that small-stakes risk aversion results from the same self-control problem that leads to a present bias in the timing of rewards. They also discuss the sizable literature that examines the correlation between cognitive ability and present bias without discussing risk aversion.

⁴⁶ We thank the authors for providing us with this data. There is data on a risky alternative involving four other size prizes that are not relevant for our purposes. There is one anomaly in the data that we cannot explain: the fraction of people choosing the risky option against the sure alternative under cognitive load actually decreases as the size of the prize is increased. This may be due to sampling error.

TABLE A3—STUDENTS TAKING THE HIGH RISK OPTION

650/0 versus 250		650/0 versus 300/200	
No load (13)	Cognitive load (21)	No load (15)	Cognitive load (22)
70 percent	24 percent	73 percent	68 percent

numbers, 1 \$US = 625 pesos and the subjects average weekly allowance was around 10,000 pesos from which they had to buy themselves lunch twice a week.⁴⁷

The key fact in the table is that introducing cognitive load when the alternative is safe induces many subjects to switch to the safe alternative, while there is no such reversal when the “safe” alternative is the 300/200 gamble. This is as the theory predicts. If the short-run self prefers the safe alternative to the risky one we should see the first reversal. However, the 300/200 gamble is less tempting than the sure alternative of 250, so a cognitive load that will lead to a reversal in the first scenario need not do so in the second.

REFERENCES

- Abdel-Ghany, Mohamed, Gordon E. Bivens, James P. Keeler, and William L. James.** 1983. “Windfall Income and the Permanent Income Hypothesis: New Evidence.” *Journal of Consumer Affairs*, 17(2): 262–76.
- Barberis, Nicholas, Ming Huang, and Richard H. Thaler.** 2006. “Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing.” *American Economic Review*, 96(4): 1069–90.
- Battalio, Raymond C., John H. Kagel, and Komain Jiranyakul.** 1990. “Testing between Alternative Models of Choice under Uncertainty: Some Initial Results.” *Journal of Risk and Uncertainty*, 3(1): 25–50.
- Baucells, Manel, and Franz H. Heukamp.** 2010a. “Common Ratio Using Delay.” *Theory and Decision*, 68(1–2): 149–58.
- Baucells, Manel, and Franz H. Heukamp.** 2010b. “Probability and Time Trade-off.” blog.iese.edu/heukamp/files/201%7/PTT_06_291.pdf.
- Benhabib, Jess, and Alberto Bisin.** 2005. “Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions.” *Games and Economic Behavior*, 52(2): 460–92.
- Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro.** 2006. “Who is Behavioral: Cognitive Ability and Anomalous Preferences.” faculty.chicagobooth.edu/jesse.shapiro/research/iq050506.pdf.
- Bernheim, B. Douglas, and Antonio Rangel.** 2004. “Addiction and Cue-Triggered Decision Processes.” *American Economic Review*, 94(5): 1558–90.
- Boldrin, Michele, Lawrence J. Christiano, and Jonas D. M. Fisher.** 2001. “Habit Persistence, Asset Returns, and the Business Cycle.” *American Economic Review*, 91(1): 149–66.
- Bossaerts, Peter, Charles Plott, and William R. Zame.** 2007. “Prices and Portfolio Choices in Financial Markets: Theory, Econometrics, Experiments.” *Econometrica*, 75(4): 993–1038.
- Brocas, Isabelle, and Juan D. Carrillo.** 2008. “The Brain as a Hierarchical Organization.” *American Economic Review*, 98(4): 1312–46.
- Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper.** 2010. “Risk and Rationality: Uncovering Heterogeneity in Probability Distortion.” *Econometrica*, 78(4): 1375–1412.
- Burke, Michael S., John R. Carter, Robert D. Gominiak, and Daniel F. Oehl.** 1996. “An Experimental Note on the Allais Paradox and Monetary Incentives.” *Empirical Economics*, 21(4): 617–32.
- Camerer, Colin F., and Teck-Hua Ho.** 1994. “Violations of the Betweenness Axiom and Nonlinearity in Probability.” *Journal of Risk and Uncertainty*, 8(2): 167–96.

⁴⁷ Many of them buy lunch at McDonald’s for 2,000 pesos twice a week, leaving an apparent disposable income of 6,000 pesos per week.

- Chetty, Raj, and Adam Szeidl.** 2007. "Consumption Commitments and Risk Preferences." *Quarterly Journal of Economics*, 122(2): 831–77.
- Conlisk, John.** 1989. "Three Variants on the Allais Example." *American Economic Review*, 79(3): 392–407.
- Constantinides, George M.** 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, 98(3): 519–43.
- Cox, James C., Vjollca Sadiraj, Bodo Vogt, and Utteeyo Dasgupta.** 2007. "Is There a Plausible Theory for Risky Decisions?" Georgia State University Working Paper 2010–06.
- Dekel, Eddie.** 1986. "An Axiomatic Characterization of Preferences Under Uncertainty: Weakening the Independence Axiom." *Journal of Economic Theory*, 40(2): 304–18.
- Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini.** 2009. "Temptation-Driven Preferences." *Review of Economic Studies*, 76(3): 937–71.
- Frederick, Shane.** 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, 19(4): 25–42.
- Frederick, Shane, George Loewenstein, and Ted O'Donoghue.** 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351–401.
- Fudenberg, Drew, and David K. Levine.** 2006. "A Dual-Self Model of Impulse Control." *American Economic Review*, 96(5): 1449–76.
- Fudenberg, Drew, and David K. Levine.** 2010. "Timing and Self Control." levine.sscnet.ucla.edu/papers/randomly_lived.pdf.
- Gabaix, Xavier, and David Laibson.** 2001. "The 6D Bias and the Equity-Premium Puzzle." In *NBER Macroeconomics Annual*, ed. Ben S. Bernanke and Kenneth Rogoff, 257–330. Cambridge, MA: MIT Press.
- Gailliot, Matthew T., Roy F. Baumeister, C. Nathan DeWall, Jon K. Maner, E. Ashby Plant, Dianne M. Tice, and Lauren E. Brewer.** 2007. "Self-Control Relies on Glucose as a Limited Energy Source: Willpower Is More Than a Metaphor." *Journal of Personality and Social Psychology*, 92(2): 325–36.
- Green, Leonard, and Joel Myerson.** 1996. "Exponential Versus Hyperbolic Discounting of Delayed Outcomes: Risk and Waiting Time." *American Zoologist*, 36: 496–505.
- Grossman, Sanford J., and Guy Laroque.** 1990. "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods." *Econometrica*, 58(1): 25–51.
- Gul, Faruk, and Wolfgang Pesendorfer.** 2001. "Temptation and Self-Control." *Econometrica*, 69(6): 1403–35.
- Gul, Faruk, and Wolfgang Pesendorfer.** 2008. "The Case for Mindless Economics." In *The Foundations of Positive and Normative Economics: A Handbook*, ed. Andrew Caplin and Andrew Schotter. New York: Oxford University Press.
- Halevy, Yoram.** 2008. "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect." *American Economic Review*, 98(3): 1145–62.
- Harrison, Glenn W.** 1994. "Expected Utility Theory and the Experiments." *Empirical Economics*, 19(2): 223–53.
- Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutström.** 2005. "Risk Aversion and Incentive Effects: Comment." *American Economic Review*, 95(3): 897–901.
- Holt, Charles A., and Susan K. Laury.** 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5): 1644–55.
- Holt, Charles A., and Susan K. Laury.** 2005. "Risk Aversion and Incentive Effects: New Data without Order Effects." *American Economic Review*, 95(3): 902–12.
- Imbens, Guido W., Donald B. Rubin, and Bruce I. Sacerdote.** 2001. "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players." *American Economic Review*, 91(4): 778–94.
- Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles.** 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589–1610.
- Kahneman, Daniel, and Amos Tversky.** 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–92.
- Keren, Gideon, and Peter Roelofsma.** 1995. "Immediacy and Certainty in Intertemporal Choice." *Organizational Behavior and Human Decision Processes*, 63(3): 287–97.
- Kőszegi, Botond, and Matthew Rabin.** 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133–65.
- Kőszegi, Botond, and Matthew Rabin.** 2009. "Reference-Dependent Consumption Plans." *American Economic Review*, 99(3): 909–36.
- Laibson, David.** 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2): 443–77.

- Landsberger, Michael.** 1966. "Windfall Income and Consumption: Comment." *American Economic Review*, 56(3): 534–40.
- Loewenstein, George, and Ted O'Donoghue.** 2004. "Animal Spirits: Affective and Deliberative Processes in Economic Behavior." Social Science Research Network Working Paper 539843.
- Marquis, Milt.** 2002. "What's Behind the Low U.S. Personal Saving Rate?" Federal Reserve Bank of San Francisco (FRBSF) Economic Letter 2002–09.
- Mehra, Rajnish.** 2003. "The Equity Premium: Why Is It a Puzzle?" National Bureau of Economic Research Working Paper 9512.
- Muraven, Mark, and Roy F. Baumeister.** 2000. "Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?" *Psychological Bulletin*, 126(2): 247–59.
- Muraven, Mark, Dianne M. Tice, and Roy F. Baumeister.** 1998. "Self-Control as a Limited Resource: Regulatory Depletion Patterns." *Journal of Personality and Social Psychology*, 74(3): 774–89.
- Neilson, William, and Jill Stowe.** 2002. "A Further Examination of Cumulative Prospect Theory Parameterizations." *Journal of Risk and Uncertainty*, 24(1): 31–46.
- Noor, Jawwad.** 2007. "Commitment and Self-Control." *Journal of Economic Theory*, 135(1): 1–34.
- Noor, Jawwad.** 2011. "Temptation and Revealed Preference." *Econometrica*, 79(2): 601–44.
- Noor, Jawwad, and Norio Takeoka.** 2010a. "Uphill Self-Control." *Theoretical Economics*, 5(2): 127–58.
- Noor, Jawwad, and Norio Takeoka.** 2010b. "Temptation by Menus." Unpublished.
- Ozdenoren, Emre, Steven W. Salant, and Dan Silverman.** Forthcoming. "Willpower and the Optimal Control of Visceral Urges." *Journal of the European Economic Association*.
- Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland.** 2010. "Consumer Spending and the Economic Stimulus Payments of 2008." www.kellogg.northwestern.edu/faculty/parker/htm/research/PSJM2010.pdf.
- Postlewaite, Andrew, Larry Samuelson, and Dan Silverman.** 2008. "Consumption Commitments and Employment Contracts." *Review of Economic Studies*, 75(2): 559–78.
- Prelec, Drazen.** 1998. "The Probability Weighting Function." *Econometrica*, 66(3): 497–527.
- Rabin, Matthew.** 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.
- Samuelson, Paul A.** 1963. "Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia*, 98: 108–13.
- Shefrin, Hersh M., and Richard H. Thaler.** 1988. "The Behavioral Life-Cycle Hypothesis." *Economic Inquiry*, 26(4): 609–43.
- Shiller, Robert J.** 1989. *Market Volatility*. Cambridge: MIT Press.
- Shiv, Baba, and Alexander Fedorikhin.** 1999. "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making." *Journal of Consumer Research*, 26(3): 278–92.
- Sydnor, Justin.** 2009. "Sweating the Small Stuff: The Demand for Low Deductibles in Homeowners Insurance." Unpublished. <http://faculty.chicagobooth.edu/workshops/AppliedEcon/archive/pdf/May%2020%20Justin%20Sydnor.pdf>.
- Thaler, Richard.** 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1(1): 39–60.
- Tversky, Amos, and Daniel Kahneman.** 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.
- Ward, Andrew, and Traci Mann.** 2000. "Don't Mind If I Do: Disinhibited Eating Under Cognitive Load." *Journal of Personality and Social Psychology*, 78(4): 753–63.
- Weber, Bethany J., and Gretchen B. Chapman.** 2005. "The Combined Effects of Risk and Time on Choice: Does Uncertainty Eliminate the Immediacy Effect? Does Delay Eliminate the Certainty Effect?" *Organizational Behavior and Human Decision Processes*, 96(2): 104–18.
- Wu, George, and Richard Gonzalez.** 1996. "Curvature of the Probability Weighting Function." *Management Science*, 42(12): 1676–90.