Title

Heterogeneity in Time Preference : : Measurement, Manipulation and Importance for Economic Policy

Author

Kuhn, Michael Anton

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Heterogeneity in Time Preference: Measurement, Manipulation and Importance for Economic Policy

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Economics

by

Michael Anton Kuhn

Committee in charge:

Professor James Andreoni, Chair
Professor Julie Cullen
Professor Gordon Dahl
Professor Uri Gneezy
Professor Yuval Rottenstreich

2014

The Dissertation of Michael Anton Kuhn is approved, and it is acceptable in quality and form for publication on microfilm an electronically.

Chair

University of California, San Diego

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I offer my sincere thanks to Professor James Andreoni for his role as an academic, professional and personal mentor throughout the authorship of this dissertation. Additionally, without the support of Nicolle Durand, Barbara Kuhn, Peter Kuhn and Jonathan Kuhn, this work would not have been possible.

Chapter 1, in full, is currently being prepared for submission for publication of the material.

Chapter 2, in full, has been submitted for publication of the material at the *American Economic Journal: Microeconomics*. Andreoni, James; Kuhn, Michael A.; Sprenger, Charles. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Kuhn, Michael A.; Kuhn, Peter; Villeval, Marie Claire. The dissertation author was the primary investigator and author of this material.

VITA

2009	Bachelor of Arts, University of California, Los Angeles
2014	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Heterogeneity in Time Preference: Measurement, Manipulation and Importance for Economic Policy

by

Michael Anton Kuhn

Doctor of Philosophy in Economics

University of California, San Diego, 2014

Professor James Andreoni, Chair

Time preferences, the willingness of decision makers to substitute intertemporally, are not shared or consistent values. They vary across populations, households, individuals and mindsets. Understanding the systematic variability of time preferences is crucial for designing effective economic policy, especially in light of the modern focus on non-standard models of discounting. This dissertation features a real-world data project that highlights the important interaction between individual differences in preferences and economic policy, a laboratory experiment that showcases the sensitivity of an individual's time preferences to environmental factors and a laboratory experiment that validates a field-ready technique for estimating an individual's time preferences that is already being implemented by many other researchers.

CHAPTER 1

Curing the Calorie Crunch: The Effect of EBT on Household Present Bias

Abstract:

Intra-month cycles in household consumption and expenditures are often considered hallmarks of dynamically-inconsistent planning. I find that food-stamp households with more children and gender-balanced adult populations experience this phenomenon more severely prior to the introduction of Electronic Benefit Transfer (EBT), and that the introduction of EBT eliminates this gap. I propose an explanation based on collective dynamic inconsistency –present bias generated by the aggregation of differing preferences within households– and that EBT affects this aggregation process by establishing more dictatorial control over the food stamp resources. This implies that policies that improve the property rights of a single recipient over a transfer disbursement and shortening the frequency of disbursement would help improve budgeting, especially for families with young children.

1.1 Introduction

Following the recent economic crises, food security in the United States is at the forefront of welfare policy. The number of individuals classified as "food insecure" by the USDA[1] grew by 14 million (roughly 28%) from 2007 to 2011, with food stamp enrollment nearly doubling from 26 to 46 million.[2] Food Stamps, now known offically as the Supplementary Nutrition Assistance Program or SNAP, is the largest government program dedicated specifically to this issue. As of 2012, one in eleven Americans received SNAP benefits in any given month.[3] Economists have highlighted the program's impact on child health (see Almond, Hoynes, and Schanzenbach (2011) for a review of this literature), food insecurity (Bhattacharya and Currie, 2000) and many types of nutritional intake (Devaney and Moffitt, 1991).[4]

There is little definitive work on how the introduction of Electronic Benefit Transfer (EBT) has affected SNAP.[5] Previous efforts analyze data on the same monthly frequency as the transfer distribution, or longer. Given the nature of food necessities, it may be that the effects of EBT need to be examined on a daily basis. I do so, using the program implementation as a tool for testing testing a theory of non-unitary household behavior and

[1]Prior to 2006, the USDA's blanket definition of food insecurity was, "consistent access to adequate food is limited by a lack of money and other resources at times during the year." In 2006, they introduced two levels of food insecurity, low food security ("reports of reduced quality, variety, or desirability of diet. Little or no indication of reduced food intake") and very low food security ("Reports of multiple indications of disrupted eating patterns and reduced food intake"). Classifications are made yearly, at the household level: http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us.

[2]IHRC (2013), p. 8.

[3]IHRC (2013), p. 12

[4]For a more comprehensive review of food assistance programs generally, see Currie (2003).

[5]Currie and Grogger (2001) and Kaushal and Gao (2011) find some positive effects on enrollment. Atasoy, Mills, and Parmeter (2010) find that it decreased enrollment. Bednar (2011) finds no significant effect.

its affect on time preference. As a motivating anecdote, consider this quotation reported in a 2013 International Human Rights Clinic study. It features Tiffany, a mother of three on food stamps with no labor income due to disability:

> My food stamps are depleted after maybe two and a half weeks. That's when our cupboards become bare and there isn't anything left in the deep freezer. I start to worry about where our next meal is coming from. (IHRC (2013) p. 20)

This downward-sloping intra-month trend in food expenditures and consumption is termed the "calorie crunch". Reconciling it with a dynamic model of consumer behavior is troubling, yet it is observed frequently.[6] Why not budget the food stamps to achieve a steadier, but lower level of consumption throughout the entire month? Shapiro (2005) attributes this behavior to present-biased time preferences; decision makers over consume in the present, without internalizing that they will fail to resist the temptation to do so again in the future. More severe present bias means a more severe calorie crunch.

Groups can exhibit present bias even if none of the individuals within the group are, on their own, present-biased. There are a variety of mechanisms through which this "collective present bias" can occur, and I focus on the framework of Jackson and Yariv (2012a). They show that under very general conditions, heterogeneity in exponential discount rates generates collective present bias when preferences are aggregated. For example, parents and toddlers have very different time preferences. Similarly, aggregation through household bargaining can lead to intertemporal trades or resource scrambles that manifest as present-biased representative preferences. While they are unlikely to bargain directly, the

[6]Stephens Jr. (2003) and Mastrobuoni and Weinberg (2009) with Social Security disbursements, Stephens Jr. (2006) with paycheck receipt, Shapiro (2005) and Hastings and Washington (2010) with Food Stamp transfers.

parent is surely aware of their child's impatience. Thus, the temptation to appease that impatience is a feature of the parent's household time preference even if it isn't a part of their personal time preference. Another approach, due to Hertzberg (2012), shows that under similarly general conditions, groups with similar discounting but different weights over the independent consumption streams within the household experience a dynamic common-pool problem that is best represented by hyperbolic-style present bias. These models have testable implications for which types of households should be subject to the strongest bias, and how this comparative static should be affected by shocks to household property rights over resources. EBT, I argue, is such a shock. By unifying control over the food-stamp resources in a variety of ways, the EBT card it could reduce household present bias to the degree that it was generated through collective means.

Specifically, I explore across-household heterogeneity in the severity of the calorie crunch and determining the effect of the introduction of EBT on the degree of this heterogeneity. This is a novel approach to analyzing this policy with direct implications for other welfare and transfer programs. I consider the consistency of the results with Jackson and Yariv (2012a) and others, providing the first direct test of the theory outside the laboratory. Additionally, I use the results (both reduced-form and structural) to asses how much empirically observed present-bias can be attributed to collective rather than individual sources: a contribution to the broader literature on dynamic inconsistency.

The results indicate substantial heterogeneity in the severity of the calorie crunch prior to the introduction of EBT, in a manner consistent with collective present bias. This finding is clearer in expenditure than consumption data. Moreover, the introduction of EBT has the predicted effect. Families with young children and gender-balanced adult populations, which exhibit the most severe pre-EBT crunches, experience the greatest reductions in severity due to EBT. The overall effect of the policy is to substantially reduce heterogeneity in crunch severity.

I compare the magnitude of the effect of EBT to other food-subsidy policies that have been shown to have positive effects on child health. While post-EBT data still exhibit a pattern consistent with meaningful individual present-bias, the portion of the pre-EBT trends attributable to collective present bias is substantial. I show that EBT affects preferences over goods as well as time, consistent with the hypothesis that its introduction has changed the preference expression balance. Additionally, I consider empirical specifications that cast doubt on the role that stigma or fraud could play in explaining the primary results.

Section 1.2 discusses the foundational empirical work in labor economics and relevant theory. Section 1.3 presents the results and Section 1.4 concludes.

1.2 Empirical and Theoretical Foundations

From a daily perspective, the majority of income receipts by households are lumpy. Paychecks and unemployment insurance usually arrive biweekly, food stamps (the largest non-elderly safety net program, by expenses, as of 2010) and disability benefits arrive monthly, and the Earned Income Tax Credit (the largest cash-transfer program, by expenses, as of 2010, for non-elderly individuals) arrives on a yearly frequency (Bitler and Hoynes, 2013). It is crucial for the evaluation of these programs to understand how income is meted out by recipients over the relevant interval, especially if recipients are likely to be cash-constrained.

Stephens Jr. (2003, 2006) demonstrates jumps in non-durable expenditures in response to social security payments and paychecks, respectively.[7] Shapiro (2005) goes a step further by identifying the decline in daily caloric intake over the course of the month following food-stamp receipt.[8] Explorations of heterogeneous trends within this literature are generally limited to identifying effects stratified by likelihood of liquidity constraint. As expected, constrained households exhibit more severe fluctuations (Mastrobuoni and Weinberg, 2009). For this reason, dynamic responses to programs targeted at distressed households, such as SNAP, should be particularly suspect. Shapiro (2005) briefly addresses whether his findings could be generated by a scramble for resources within a household. He finds no economically or statistically significant effect of household size on cycle severity. I revisit this issue with a broader approach to classifying households as subject to detrimental bargaining processes, an exogenous shifter of weights in the bargaining process and with data that offer more statistical power.

1.2.1 How Can Household Bargaining Affect Expenditure Cycles?

Hyperbolic and quasi-hyperbolic time preferences (Laibson, 1997; O'Donoghue and Rabin, 1999) feature higher discount rates in the short run than in the long run. Harris and Laibson (2001) show that individuals solving a standard budgeting problem with

[7] Stephens Jr. (2003) uses the same data source as the present study: the Consumer Expenditure Survey. However, the limited nature of heterogeneity across recipients of social security in terms of household structure makes the group less relevant for the issue at hand.

[8] Mastrobuoni and Weinberg (2009) extends the Stephens Jr. (2003) finding to caloric intake and Hastings and Washington (2010) extends Shapiro (2005) using grocery store scanner data rather than self reports.

these preferences will not save precautionarily and exhibit a "comovement between income and consumption" (p. 937). Shapiro (2005) recognizes an empirical manifestation of this in downward-sloping intra-month consumption profiles. A contribution of the present work is to suggest that the theory of dynamic inconsistency generated by collective decision making, as explained by Jackson and Yariv (2012a) and Hertzberg (2012), can justify the use of hyperbolic preferences to explain this phenomenon without necessitating non-standard time preferences on the individual level. Additionally, these theories have direct implications for which households, all-else-equal, should be subject to collective decision processes that generate these cycles.

I illustrate this with an example based on Jackson and Yariv (2012a). Consider two individuals, a and b who are jointly evaluating a time-stamped consumption stream. They are exponential discounters with different rates of positive discounting, $\delta_a \neq \delta_b, \delta_a, \delta_b \in [0, 1]$, with preferences

$$V_a = \sum_{t=0}^{T} \delta_a^t u(c_t) \quad \text{and} \quad V_b = \sum_{t=0}^{T} \delta_b^t u(c_t). \tag{1}$$

Their preferences are equally weighted in the collective decision process such that

$$V_{a,b} = 0.5V_a + 0.5V_b = \sum_{t=0}^{T} \frac{\delta_a^t + \delta_b^t}{2} u(c_t) = \sum_{t=0}^{T} \Delta_t u(c_t), \tag{2}$$

where $V_{a,b}$ is the collective utility function and Δ_t is the collective discount factor for period

t. The first 3 discount factors are

$$\Delta_0 = 1, \quad \Delta_1 = \frac{\delta_a + \delta_b}{2} \quad \text{and} \quad \Delta_2 = \frac{\delta_a^2 + \delta_b^2}{2}. \tag{3}$$

What determines the group's relative preferences for one period over another is the ratio of the corresponding discount factors. For example,

$$\frac{\Delta_0}{\Delta_1} = \frac{2}{\delta_a + \delta_b} \geq \frac{\Delta_1}{\Delta_2} = \frac{\delta_a + \delta_b}{\delta_a^2 + \delta_b^2}, \tag{4}$$

meaning that relative preferences for consumption in adjacent periods is non constant. Specifically, the relative preference for consumption sooner is greater when the sooner period is closer to the present, holding the gap between the sooner and later periods fixed.[9] In other words, the collective preferences are present-biased. Most generally, Jackson and Yariv (2012a) prove that any non-dictatorial, unanimity respecting aggregation rule will generate group preferences that are dynamically inconsistent, provided differing constituent discount factors.[10]

Bigger differences in preferences will generate more dynamic inconsistency and thus bigger budget shortfalls. If the exponential discount factors, δ_a and δ_b are very similar, the degree of collective present bias will be very small when the consumption stream is common.[11] There is another avenue by which collective present bias is generated, however.

[9]This follows algebraically: $\frac{2}{\delta_a+\delta_b} \geq \frac{\delta_a+\delta_b}{\delta_a^2+\delta_b^2} \Rightarrow 2(\delta_a^2+\delta_b^2) \geq \delta_a^2+\delta_b^2+2\delta_a\delta_b \Rightarrow \delta_a^2+\delta_b^2-2\delta_a\delta_b \geq 0 \Rightarrow (\delta_a-\delta_b)^2 \geq 0$, which must be true.

[10]Others, including Hertzberg (2012), Zuber (2011) and Bernheim (1999) allude to similar effects of preference aggregation under different circumstances, most notably, separate consumption streams and identical discount factors.

[11]Specific to models of household preferences in which the preferences of individuals are additively sep-

Hertzberg (2012) proves that when consumption streams are at least partly independent and individuals within the household weight the separate streams differently (for example, weight their own consumption higher than that of others), even individuals with identical exponential discount factors generate households with representative preferences that are hyperbolic through bargaining. Altogether, the finding of non-stationary group discounting is highly general.

1.2.2 Evidence on Individual Time Preference in a Low-income Sample

I take advantage of a unique dataset collected by Andreoni et al. (2011) from the Griffin Early Childhood Center (GECC) in Chicago Heights, Illinois. The area is a low-income suburb of Chicago, and the preschool services families that would be of interest for food stamp policy.

Preschoolers' time preferences were elicited by having them make a series of choices between two plates of candy. One plate could be received on the current day and the other could be received the next day. Four candies were always on the "today" plate, while the number on the "tomorrow" plate ranged from five to eight as the choices progressed. 75% of children chose the "today" plate when the opportunity cost was 1.25 candies tomorrow, and this only decreased to 61% as the opportunity cost reached 2 candies tomorrow. The data showcase extraordinary levels of discounting that most parents would find completely

arable, the consumption stream of the group is bounded between the streams that would be generated by individuals decision making. This is why this simple version of the model is most useful for thinking about the incorporation of the preferences of indivdiuals who exert influence on others in the household, but are less capable of stealing or forcibly acquiring resources for themselves (like children). Models of intertemporal bargaining or commons problems with non-separable solutions generate group behavior beyond the bounds of individual behavior, are are thus more appropriate for considering interactions between adults.

unsurprising. Additionally, the well-known marshmallow experiment conducted by Mischel, Ebbesen, and Raskoff Zeiss (1972) finds that 3-5 year olds resiting temptation in front of potential rewards (in my context, in a supermarket) last less than 30 seconds when attempting to delay gratification to double their reward. More recently, Levit et al. (2012), shows that delayed rewards (in contrast to immediate rewards) have no motivational effect on primary and secondary school students. Bettinger and Slonim (2007) shows the same with direct measures of time preference, but also shows that discounting converges quickly to adult levels during adolescence.

The effect of highly impatient individuals on household time preferences can be very large. Consider a household consisting of a parent (the decision maker) and their infinitely impatient ($\delta_c = 0$) child. The parent has standard exponential preferences with discount factor, δ_h, for the household as a whole[12] and the child has preferences for itself alone. If household preferences are a weighted sum of the two sets, with weight B assigned by the decision maker to their own preferences, two-period preferences for the household, $V(\cdot)$, are

$$V(c_t, c_{t+1}) = \begin{cases} u(c_t) + B\delta u(c_{t+k}) & \text{if } t = 0 \\ u(c_t) + \delta u(c_{t+k}) & \text{if } t > 0 \end{cases}, \tag{5}$$

where a common factor of $(1 - B)$ has been removed from the case in which $t > 0$. This is exactly the quasi-hyperbolic model of time discounting. Adding more children has a natural interpretation of decreasing B, thus increasing the expressed present-bias. Moreover, it takes little appeasement to cause budgeting problems; a value of 0.95 for B

[12]This seems more reasonable than to suggest the parent has preferences only for themselves.

represents meaningful present-bias.

Andreoni et al. (2011) also elicited parental discounting. Their time preferences are estimated using a Convex Time Budget (Andreoni and Sprenger, 2012) in the specific form of Andreoni, Kuhn, and Sprenger (2013). At least over monetary disbursements, the average individual does not exhibit present-biased preferences[13] and discounts at about 92% over a four-week period.[14] While the annualized rate associated with this estimate is very large by some standards, an interest rate of about 180% is actually below some payday loan rates and other credit relevant for this population. Importantly, women and men exhibit different short-run and long-run discounting in this sample.[15]

A key factor that determines whether gender differences in preferences lead to collective present bias is whether decisions are made jointly. Parents in the Andreoni et al. (2011) sample were asked, “Thinking back over the past month, how involved were you in your family’s financial decisions?” Answering that they made about half of the decisions (indicating multiple decision makers) is correlated with poor spending plan development, shifting the median report from “I made a spending plan and followed it some of the time” to “I have not made a spending plan in the past month” ($p < 0.05$).

Given this evidence, I explore not just the immediate implication of Jackson and Yariv (2012a), that households consisting of multiple individuals are exposed to the potential for inconsistency via preference aggregation while unitary households are not, but also

[13]There is a discussion in the literature as to whether present-bias should be observed over receipts of monetary income. Augenblick, Niederle, and Sprenger (2013) show experimentally that the same subjects who do not exhibit a disproportionate demand for income today do display a disproportionate propensity to procrastinate when scheduling labor effort if it can be allocated to the present.

[14]That is, for the utility function $U(c_t, c_{t+28}) = u(c_t) + \delta u(c_{t+28}), \delta = 0.92$.

[15]Men discount significantly higher in the short run, but significantly lower in the long run.

direct predictors of collective present bias: adding individuals to a group that have different preferences. This is why the number of children in a household will be the key variable of interest. The decision makers are sure to consider their preferences, at least to some minor degree, and their time preferences are vastly different from adults'. Gender differences in preferences only translate into dynamic inconsistency when one group cannot overrule the other.[16] Therefore, I classify households based on whether their adults are gender-balanced, and thus have equal-sized coalitions to represent the different sets of preferences.[17]

1.2.3 Empirical Application: The Introduction of EBT

Without a shock to the bargaining process, it would be difficult to draw a causal link between household composition and the severity of decline. I argue that the introduction of EBT is a shock to the bargaining process, in the form of an improvement of the property rights over the transfer for the recipient, which has the unintended consequence of changing the level of household present bias. This puts the work squarely at the intersection of behavioral economics and public finance, an area which has seen substantial growth recently.

Two long-standing issues with the Food Stamp program prior to the implementa-

[16] The language 'dictatorial' could be misleading in this context. What matters is not whether only one person is making the decisions, but whether only one set of preferences determines them. In the example above, the single parent is the solitary decision maker, but because they aggregate their preferences with their child's preferences, the result is still dynamically inconsistent.

[17] This is an approach to the data consistent with Jackson and Yariv (2012a) in which preference differentials matter, but potentially less consistent with Hertzberg (2012) in which degree of selfishness matters. A particularly interesting way to characterize households would be by income-earning structure. A number of papers, including Hoddinott and Haddad (1994) and Andreoni, Brown, and Rischall (2003) identify relative income as an important determinant of whose preferences are expressed in expenditure data. This would get at the issue of income property rights influencing bargaining weights in the aggregation process. However, the diversity in earnings structure is very low in the CEX sample; only 12% of household have 2 or more earners.

tion of Electronic Benefit Transfer were the stigma associated with the use of the visually identifiable stamps and the lack of property rights that allowed them to disperse or be sold. The EBT cards were designed with these issues in mind. They work and look like standard debit cards (see Figure A1.1 in the appendix), such that a Personal Identification Number (PIN) is needed to use them. Food stamp benefits are loaded onto the cardholder's account on a monthly frequency, but the disbursement dates vary by state.[18] When comparing this to the cash-similar coupons that were in place before (see Figure A1.2 in the appendix), I argue that the recipient is more clearly delineated as the owner and controller of the benefits following the policy change.[19]

What could this delineation of ownership mean for the households of interest? In the case of gender balance, it is clear that the property rights of the direct recipient will be improved. This works both explicitly via the choosing of a PIN number, and implicitly via the codification of the food stamps as belonging to one person. Households in some states can have more than one card or authorize additional users, *but only with the permission of the original recipient*. If this constitutes a shock in the direction of dictatorial decision making, EBT should counteract the calorie crunch for these households. Considering the number of children, if they are old enough to go to a store themselves, the coupons could be disbursed more easily than the EBT card, which makes it more likely that the card holder is making purchasing decisions. For younger (but not too young) children, EBT gives the

[18]Some disburse on the first of the month, but many disburse on different days based on last names, EBT identification numbers or social security numbers.

[19]The laws vary from state to state about who can use the cards, and whether the cards feature some individual-specific identifiers like a photograph. For example, in New York the cards feature a photo, name, gender, birthdate and a signature whereas the Illinois card has none of these features. The common element is a PIN number of the primary recipient's choosing that they can keep secret and change.

card holder the ability to lie about the amount of benefits that remain because they aren't visible. The more formal property rights implied by the card could also simply reinforce mom or dad's willpower in saying no to the kids.

A broad-strokes outline of the the hypothesis I test is that *the less homogeneous the constituent preferences of a household are and the less dictatorial the aggregation process is, the more present-biased household preferences will be, which will translate into more severe expenditure and consumption cycles. Policies that allow the transfer recipient to exert more dictatorial control over income will work against this heterogeneity.*

This paper is relevant to a larger literature in labor economics on non-unitary household behavior and transfer efficacy. A well-known example is given by Lundberg, Pollak, and Wales (1997), who observe that shifting child allowance transfers to wives in the U.K. increased household expenses on women's (and children's) clothing.[20] While recipient effects are important for all sources of income, it is of particular importance for transfer income because of the associated policy objectives. Food stamps are, ideally, used to help provide low-income families with a *steady* flow of *nutritious* food. Studies that examine purchases on the same frequency as the transfer miss out on the first of the two goals. If intra-household preference aggregation can help explain both the stylized facts of intra-month expenditure patterns and the types of items demanded, then there is a clear advantage to dealing with both issues simultaneously.

There is some experimental work that analyzes the relationship between group pref-

[20]Especially in the low-income and developing context, and number of studies indicate differing expenditures between men and women in the same household, depending on income recipient. Examples include Browning et al. (1994), Browning and Chiappori (1999), Duflo (2003), Bobonis (2009), Attanasio and Lechene (2011), Attanasio, Battistin, and Mesnard (2012) and Wang (2012). It is straightforward to embed a "wallet-to-purse" model within the collective present bias explanation of my results.

erences and the preferences of the constituents. Results are mixed. Abdellaoui, l'Haridon, and Paraschiv (2013) find that joint discounting decisions of real couples cannot be explained as a mix of the individual preferences. However, Carlsson et al. (2013) find that for risk preferences, couples decisions end up somewhere in the range between individual decisions. Additionally, Jackson and Yariv (2012b) find that subjects in the lab acting as social planners exhibit considerable time inconsistency. This paper is, to the author's knowledge, the first test of collective present bias outside of the lab.

1.3 Results

I build up to examining the effect of the EBT introduction in three parts. First, I replicate previous literature by demonstrating a significant and robust decline in food expenditure and consumption. Second, I identify heterogeneity in the severity of this decline across households of different types prior to EBT. Third, I show that EBT significantly reduces (and often eliminates) the heterogeneity. At the end of the section, I use a structural model to characterize the magnitude of the heterogeneity and policy effects in the context of quasi-hyperbolic discounting and explore an alternative explanation for the results.

1.3.1 Intra-month Expenditure Profiles

The first data I use to identify downward-sloping intra-month expenditure profiles are the Consumer Expenditure Survey (CEX) Diaries. These are self-reported expenditure logs that cover 14 consecutive days. It is collected every year, all throughout the year. The diaries consist of two back-to-back, week-long logs formatted for respondents to keep item-specific records of all purchases. These diaries are linked to a broader, one-time survey of household demographics, composition and income. Food purchases are separated from other purchases. Following collection, the item-level data are coded with a Universal Classification Code (UCC), which permits the identification of expenditures on food alone. Purchases are not coded as individual-specific.

The usable set of CEX diaries ranges from 1994 to 2003. Prior to and following this sample period, the CEX did not record the exact date of the most recent food stamp arrival. Conditioning on households with recorded expenditures within 4 weeks of their recorded food stamp disbursement yields 1302 households over 14,809 days.[21]

I follow the panel approach of Shapiro (2005) with households i and 14 diary days j. Given the date of most recent food stamp receipt, j is transformed into a variable that indicates the number of days since then t, with $t = 0$ on the exact date of reported arrival. If $t = 0$ corresponds with $j = 1$ for a particular household, all 14 days of the diary are used as the first 2 weeks of expenditures. So long as $j = 1$ corresponds to $0 \leq t \leq 14$, all 14 diary days are potential shopping days. I do not use diary observations that fall outside of the food stamp period corresponding to the reported receipt because of the uncertainty as to whether the disbursement was identical the month prior to or following the reported one. This setup implies the basic fixed-effect specification for expenditure trends,

$$exp_{it} = \alpha_i + \beta t + \Gamma Y_t + \epsilon_{it}, \tag{6}$$

[21]FS eligible purchases are made on slightly less than 35% of days in the sample. Households consist of an average of 3.2 members, with 1.5 of them younger than 18 and 1.1 of them under 13.

where exp_{it} are the food expenditures of household i on days since food stamp receipt t and Y_t are other characteristics of the day in question to be controlled for: a weekend dummy, a week of month variable and a week of diary variable. The identifying assumption for equation (6) is that day-to-day *changes* in expenditures are conditionally uncorrelated with the error term. This might fail if different types of households inform different parts of the food-stamp month. I establish in appendix Table A1.1 that important observables are uncorrelated with when, in the food-stamp month, households are observed.

Table 1.1 presents a variety of specifications that correspond to different sample restrictions and periods of observation. I strongly corroborate the existence of expenditure cycles with downward sloping intra-month profiles. The most basic specification in column (1) estimates a decline of $1.85 per day including all potential shopping days, dropping to $1.52 or roughly 6% per day conditional on non-trivial expenditures in columns (2) and (3).[22] Collapsing the data by week since arrival of food stamps reduces the gap between the conditional and unconditional samples, yielding an estimate of $13.15 or roughly 18% per week. Both the finding that most of the effect has to do with amount purchased rather than frequency of purchase and the 18% per week estimate correspond closely to the estimate Hastings and Washington (2010) obtain from grocery store scanner data.

While there are a variety of ways to model this trend, the raw data speak for themselves in this case. Figure 1.1 shows the raw data, conditional on expenditures of more than $1. Mean daily expenditures, conditional on the restriction, fall from $54.88 to $17.24 and unconditional expenditures from $27.30 to $5.77 over the month. Figure 1.1 also demon-

[22]This implies that the use of the log specification to approximate percent changes is only valid for a couple days at the beginning of the month. This will occur repeatedly throughout this section.

Table 1.1: Expenses on Food Decrease throughout the Food-Stamp Month

Dep. Var.	\$ / Day		ln(\$) / Day	\$ / Week	ln(\$) / Week
	(1)	(2)	(3)	(4)	(5)
Periods since receipt	-1.851*** (0.122)	-1.520*** (0.247)	-0.060*** (0.009)	-13.147*** (1.213)	-0.176*** (0.041)
Expenses $\geq$ \$1 Only	N	Y	Y	N	Y
Observations	14,809	5000	5000	2918	2285
Clusters	1302	1296	1296	1302	1296

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effects at the household level. Columns (1)-(3) have controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. Household weights change from month to month in the sample, so the specifications presented above are using unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results as do OLS or random-effect specifications that utilize the changing weights properly.

strates that the negative slope of the fitted trend (using OLS this time for simplicity) remains negative and significant even when the large peaks on days 0 and 1 are removed. Extending this exercise indicates that the conditional decrease drops only from 6% per day to 5.1% per day when the entire first week is removed.

Daily expenditure declines do not translate one-for-one into consumption declines. It is easier to stockpile certain types of food for a month than calories. If there were no trend for perishables, these trends in expenditure mean little for trends in consumption. However, replacing all food expenditures with perishable expenditures[23] in each specification still yields negative and significant trends. Importantly, this is true for the weekly specification, which won't predict an overall monthly decline if there exists only within-week declines based on shopping frequency. Results of this exercise are in appendix Table A1.2.

I use the Continuing Survey of Food Intake by Individuals (CSFII) as does Shapiro

[23] I use fruit, vegetables and milk, taking a narrow definition to ensure perishability.

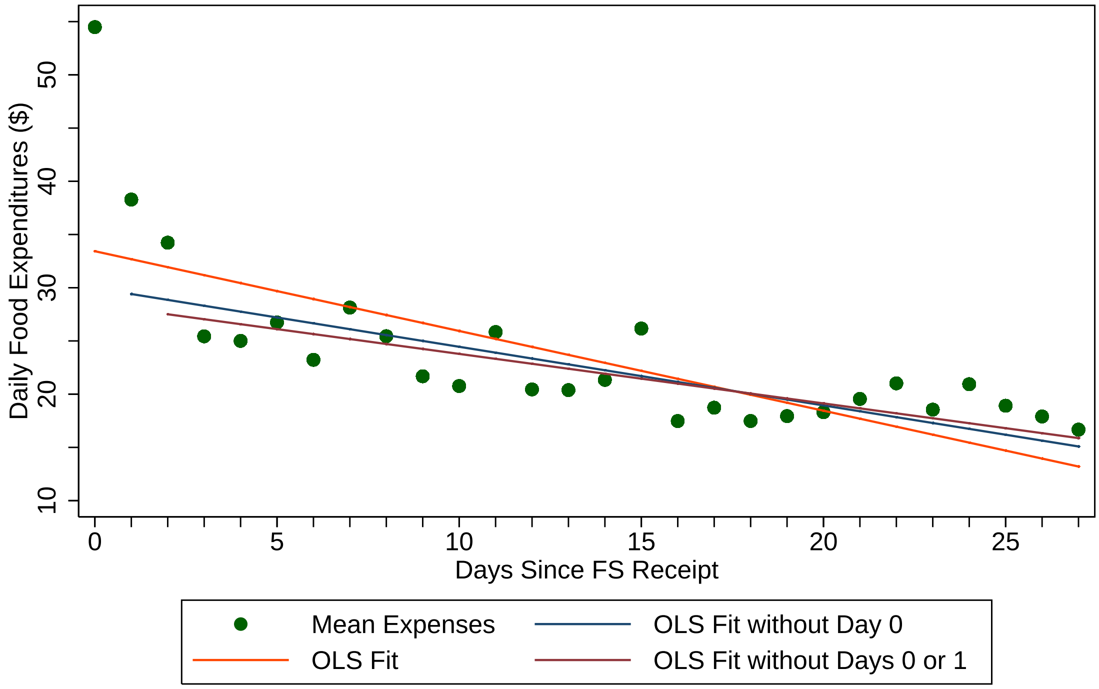


Figure 1.1: Raw Data Exhibits Strong Decline, Robust to Removal of Early Days

(2005), to illustrate the correspondence between consumption and expenditure. The CSFII is not collected annually, and the most recent version, from 1998, does not record the exact date that food stamps arrive to a household. The 1989-1992 collection does include this information, as well as demographic data. The dependent variable of food expenditures is replaced with total calories consumed by a household on a survey day. Unfortunately, the CSFII only offers up to 3 days of observations for any given household. I follow Shapiro (2005) by including dummies for survey day (1-3), day of week (1-7), month, year and calendar date rather than a fixed effect, because of the short length of each household's panel.[24] The identification assumption without the fixed effect is that the level of caloric intake is conditionally uncorrelated with the error term.

[24]There is no theoretical reason why a fixed effect cannot be used. Results are even stronger in specifications that match my CEX approach exactly.

While the data contain a report of the household size, there is no information on the ages of individuals that do not contribute diaries. Therefore, I restrict my attention to households with a consistent age profile, as represented by the individuals with diaries. I do this to cut down on measurement error associated with incorrect characterizations of households.[25] I follow my convention of only including observations occur in the 4 weeks following a reported food-stamp disbursement, although my results are robust to defining the days since receipt variable less conservatively. This leaves me with 758 households and 1861 household-days. I estimate a decline of roughly 27 kCal per-day ($S.E. = 14.43$), per-household, or about 0.7% ($S.E. = 0.4\%$) in the log specification.

1.3.2 Pre-EBT Heterogeneity in the Severity of Intra-month Declines

To establish differential trends across households, I add interaction terms between household composition variables and the number of days since food stamp receipt to the regression specifications from the previous section. While these estimates will not establish a causal link between household structure and consumption or expenditure patterns (because the compositional variables are related to a host of unobservables) they provide context for the treatment effects of the EBT program that follow in the next section.

In Section 1.2, I identify household size, number of children and gender balance as variables that a preference aggregation theory would predict are correlated with an exacerbated decline in intra-month expenditure and consumption. I first interact household size with the number of days since food stamp receipt and add this to the specification in equa-

[25] Given that incorrect characterization is more likely in larger households with potentially different expenditure profiles, this measurement error is unlikely to be classical.

Table 1.2: Household Size Not A Predictor of Pre-EBT Profile Slope

	Expenditures		Consumption	
	\$ / Day (1)	ln(\$) / Day (2)	kCal / Day (3)	ln(kCal) / Day (4)
Days since receipt	-1.494*** (0.319)	-0.082*** (0.013)	-43.133 (35.680)	-0.004 (0.008)
Days since receipt X HH Size	-0.111 (0.085)	0.004 (0.003)	8.971 (10.673)	0.001 (0.002)
Observations	3010	3010	1861	1861
Clusters	745	745	758	758

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. Expenditure models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. The household size variable is adjusted by subtracting 1, such that the terms non interacted with size apply to single-individual units. Expenses less than \$1 are trimmed to reduce noise, especially important in the log specifications. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results. Consumption models control for the amount of the food stamp disbursement and the level effect of household size.

tion (6). I limit the sample to non-trivial expenditures to focus on value of food purchased rather than frequency of food purchase. Results are presented in Table 1.2. There is no evidence that profiles were steeper for larger households.

Next, I consider different types of individuals in a household as a more direct proxy for variance in preferences. Table 1.3 presents the results from estimating profile heterogeneity by number of children, conditional on family size. I break children into 3 age groups: pre-speech fluency (ages 0-4), speech fluency, pre-adolescence (5-12) and adolescence (13-17). In this case, the results are more in line with the theoretical predictions. Prior to the implementation of EBT, replacing an adult member of a household with a child under the age of 13 exacerbates the downward trend in expenditures by 1.9% on top of a

Table 1.3: Pre-EBT Profiles Are Steeper with More Young Kids

	Expenditures		Consumption	
	\$ / Day (1)	ln(\$) / Day (2)	kCal / Day (3)	ln(kCal) / Day (4)
Days since receipt	-1.534*** (0.318)	-0.083*** (0.013)	-27.307 (38.244)	0.001 (0.007)
Days since receipt X Kids 0-5 Yrs.	-0.385* (0.230)	-0.016** (0.008)	-34.503 (27.943)	-0.005 (0.005)
Days since receipt X Kids 6-12 Yrs.	-0.538** (0.271)	-0.019* (0.010)	-15.767 (18.324)	-0.004 (0.004)
Days since receipt X Kids 13-17 Yrs.	0.108 (0.223)	-0.003 (0.009)	-34.208 (27.379)	-0.012** (0.005)
Observations	3002	3002	1861	1861
Clusters	737	737	758	758

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. Standard errors clustered by household. Expenditure models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from, which week of the month it comes from, the interaction between household size and days since food stamp receipt and the interactions between the number of children over the age of 12 and days since food stamp receipt. The household size variable is adjusted by subtracting 1, such that the terms non interacted with size apply to single-individual units. Expenses less than \$1 are trimmed to reduce noise, especially important in the log specifications. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results. Consumption models control for the amount of the food stamp disbursement, the level effect of household size, the interaction of size and days since receipt and the level effect of the number of children of each age type.

baseline decline of 8.3% (\$0.54 on a baseline of \$1.53 in the level model). The consumption estimates are not precise, but the magnitudes of the interaction correlations are clearly important, given the size of the baseline decline.

Moving to a measure of preference variance *and* non-dictatorial decision making, Table 1.4 presents the results from estimating heterogeneity by whether the adult population of a household is gender balanced, again conditional on family size.[26] The correlation is

[26] I do this because many households have 1 or 2 adults.

Table 1.4:Pre-EBT Profiles are Steeper with Gender Balance

	Expenditures		Consumption	
	\$ / Day (1)	ln(\$) / Day (2)	kCal / Day (3)	ln(kCal) / Day (4)
Days since receipt	-1.422*** (0.322)	-0.081*** (0.013)	-50.370 (30.686)	-0.006 (0.008)
Days since receipt X Males = Females	-0.619* (0.348)	-0.006 (0.013)	-77.743** (35.071)	-0.016** (0.008)
Observations	3002	3002	1861	1861
Clusters	737	737	758	758

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. Standard errors clustered by household. Expenditure models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from, which week of the month it comes from and the interaction between household size and days since food stamp receipt. The household size variable is adjusted by subtracting 1, such that the terms non interacted with size apply to single-individual units. Expenses less than \$1 are trimmed to reduce noise, especially important in the log specifications. Consumption models control for the amount of the food stamp disbursement, the level effect of household size, the interaction of size and days since receipt and the level effect of gender balance. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results.

strong in all specifications except for the log-expenditure model in column (2).

As stated earlier,estimates in this section cannot necessarily be interpreted as causal. The consumption estimates are especially suspect given the lack of a household fixed effect. However, if the differences across households have to do with the expression of different time preferences, there should be detectable differences in purchased bundles that stem from the expression of the corresponding different preferences over goods. This turns out to be true with the purchase of fresh vegetables. Aggregating the CEX diaries within households and taking advantage of the item-level data allows me to ask whether households with more young children purchase fresh vegetables less frequently (conditional on family size, food-stamp disbursement and year dummies) prior to EBT. Unsurprisingly, this is true. Replacing an adult with a child under 13 is correlated with about 0.2 fewer diary days in which fresh vegetables are purchased, from a baseline of 0.87 days on average. Adjusting for different diary lengths, this is roughly a decrease of 0.01 in the probability of purchasing fresh vegetables on any given day, from a baseline probability of 0.09.

1.3.3 Effect of EBT Implementation on Heterogeneity

In 1989 Maryland became the first state to begin the implementation of a statewide system, with completion occurring in April of 1993.[27] A number of states began implementing the program of their own accord until 1996, when the U.S. congress passed a welfare reform bill that mandated the full implementation of EBT across the country by October of 2002.[28] On average it took states about 15 months to institute the program. I use the month of official completion as the policy change date. All results are robust to the exclusion of the rollout period from the sample. See appendix Table A1.3 for a list of states and completion dates.

Because of data limitations, only the CEX data permits a study of the policy change. I amend the specification in equation (6) with an interaction between EBT implementation and days since food stamp receipt to establish a baseline effect of the program, an interaction between the household composition variable of interest and days since receipt to replicate the pre-EBT patterns from the previous section and a triple interaction between EBT, the household variable and days since receipt, which will be the variable of interest.

While the specifications all include a fixed effect, the identification of the policy

[27] http://www.fns.usda.gov/snap/ebt/ebt_status_report.htm

[28] A number of states were unable to comply until 2003. California and Guam did not complete implementation until 2004: http://www.fns.usda.gov/snap/rules/Legislation/about.htm.

effect is between-households. Therefore, it is important to establish stability of the sample composition across the EBT implementation. A useful feature of this particular policy change that greatly aids identification is that EBT rollout occurred at different times in different states over roughly a 10 year period. Thus, I am simultaneously comparing different cohorts within the same state and the same cohort across states, rather than one or the other.[29] Appendix Table A1.4 presents means of observable variables across the EBT implementation. It appears that families on food stamps in the CEX sample are slightly larger following the implementation of EBT. This corresponds to a small increase in the amount of the food stamp disbursement, one factor that could potentially affect the dynamics of spending. However, controlling for household size, there is no significant increase in the disbursement following the introduction of EBT. This indicates that families of similar compositions have similar food stamp endowments on either side of the policy change.

Following the structure of the previous section, I start by considering the effect of the EBT implementation on expenditure profile heterogeneity by household size. Results are in Table 1.5. Column (1) replicates the pre-EBT estimates for the limited sample of policy-switching states, and column (2) presents the corresponding post-EBT implementation estimates. Columns (3)-(5) combine the pre and post data to estimate the interaction terms of interest. There is no evidence of a significant effect of the interaction between the implementation of EBT and household size on the intra-month expenditure profile.

Moving to a direct measure of preference variance, the number of children, produces a very different result in Table 1.6. Using the estimates from the log specification in

[29] I limit the samples to states that switched during my period of observation so that state specific effects cannot be conflated with policy effects.

Table 1.5: EBT Doesn't Interact with Household Size

Food Stamp Type	Coupons	EBT		All	
	(1)	(2)	(3)	(4)	(5)
Days since receipt	-1.594*** (0.284)	-0.677 (0.432)	-1.239*** (0.136)	-1.274*** (0.306)	-0.071*** (0.013)
Days since receipt X HH Size	-0.128 (0.099)	-0.182* (0.106)	-0.128 (0.085)	-0.146 (0.098)	0.002 (0.004)
Days since receipt X EBT			0.183 (0.320)	0.087 (0.338)	0.022 (0.017)
Days since receipt X HH Size X EBT			-0.019 (0.132)	-0.000 (0.143)	-0.005 (0.005)
Policy Switch Only	Y	Y	N	Y	Y
Log Expenses	N	N	N	N	Y
Observations	2415	1881	5000	4296	4296
Clusters	610	533	1296	1141	1141

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. The household size variable is adjusted by subtracting 1, such that the terms non interacted with size apply to single-individual units. Expenses less than \$1 are trimmed to reduce noise, especially important in the log specifications. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results.

column (5), households without children experience an expenditure decline of 7% per-day. This interpretation of the log coefficient holds only for the first couple days of the month, indicating that an alternative speficiation of the model may be preferable. See appendix Table A1.5 for estimates that follow Hastings and Washington (2010) by using weekly dummies for expenditure levels. Pre-EBT, each young child correlates with an exacerbating of that decline by a little more than 2% per-day. EBT corrects that exacerbation, mitigating the slope by a little more than 3% for both of the younger age groups. EBT does not appear to

Table 1.6: EBT Erases Heterogeneity Associated with Young Children

Food Stamp Type	Coupons	EBT		All	
	(1)	(2)	(3)	(4)	(5)
Days since receipt	-1.616*** (0.355)	-0.477 (0.442)	-1.248*** (0.282)	-1.262*** (0.303)	-0.070*** (0.013)
Days since receipt X Kids 0-5 Yrs.	-0.378 (0.245)	0.446 (0.331)	-0.363 (0.229)	-0.356 (0.243)	-0.020** (0.008)
Days since receipt X Kids 6-12 Yrs,	-0.604* (0.307)	0.753** (0.337)	-0.502* (0.271)	-0.567* (0.311)	-0.020* (0.011)
Days since receipt X Kids 13-17 Yrs,	0.116 (0.256)	0.155 (0.302)	0.113 (0.225)	0.115 (0.258)	-0.001 (0.010)
Days since receipt X EBT			0.267 (0.338)	0.165 (0.359)	0.023 (0.018)
Days since receipt X Kids 0-5 Yrs. X EBT			0.640 (0.391)	0.728* (0.402)	0.032** (0.016)
Days since receipt X Kids 6-12 Yrs. X EBT			1.043** (0.412)	1.207*** (0.449)	0.031* (0.017)
Days since receipt X Kids 13-17 Yrs. X EBT			-0.050 (0.374)	0.011 (0.405)	0.020 (0.016)
Policy Switch Only	Y	Y	N	Y	Y
Log Expenses	N	N	N	N	Y
Observations	2415	1881	5000	4296	4296
Clusters	610	533	1296	1141	1141

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. Additionally, I control for household size interacted with both pre- and post-EBT trends such that the estimates refer to the composition rather than the size of the household. Expenses less than $1 are trimmed to reduce noise, especially important in the log specifications. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results.

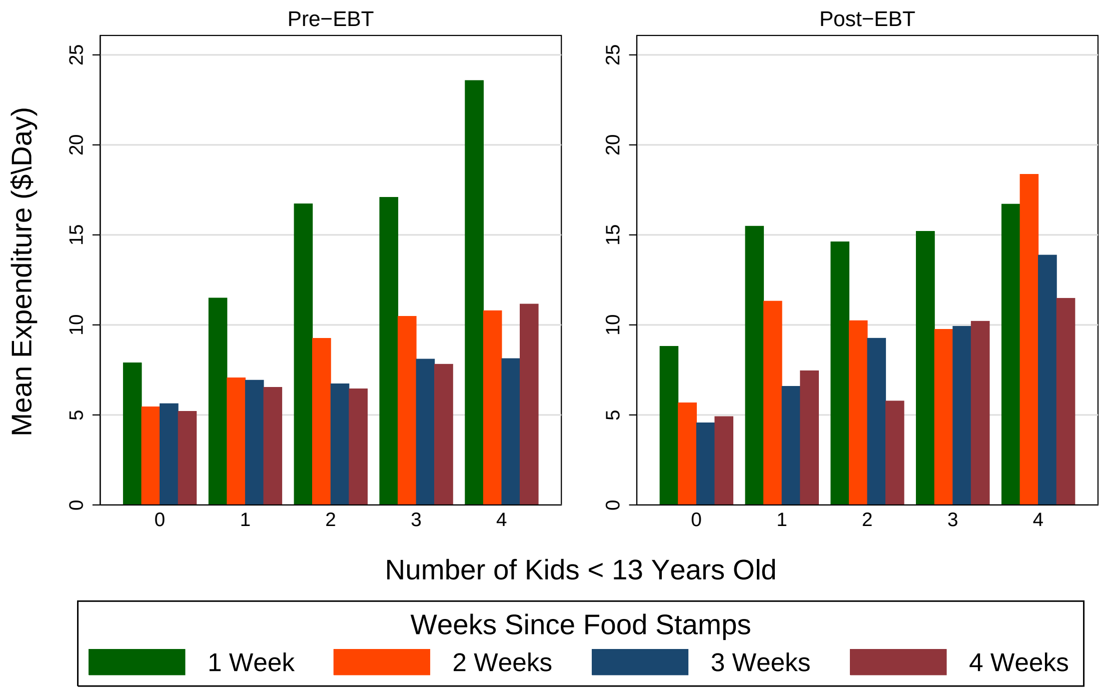


Figure 1.2: EBT Significantly Affects Relationship between Kids and Trend

have a significant baseline effect that can be applied to all households.[30]

Figure 1.2 uses just the raw data to give a visual representation of the effect of the EBT implementation for households with varying numbers of young children. I estimate pre- and post-policy change trends using OLS for each different type of household. Level differences in consumption are not absorbed, meaning that the convergence (and even overshoot) by households with many young children in Panel A is a legitimate budgeting failure that is much less serious post-EBT in Panel B.

If EBT is indeed affecting whose preferences are being expressed at the grocery store, there should be a heterogeneous effect of EBT on fresh vegetable purchasing. Con-

[30]The large change in the baseline decline estimate between columns (1) and (2), which splits the pre- and post-EBT samples appear to be at odd with the direct estimates of the EBT-days-since-recepit interaction in columns (3), (4) and (5). This could be due to the change in sample composition, but it should not affect the estimation of the triple interactions conditional on the regular interactions.

Table 1.7: EBT Increases Fresh Vegetable Purchase for Households with Children

Food Stamp Type	Coupons	EBT	All	
	(1)	(2)	(3)	(4)
Kids < 13	-0.196** (0.079)	-0.009 (0.064)	-0.174** (0.072)	-0.210*** (0.077)
Kids > 12	-0.166* (0.098)	0.047 (0.088)	-0.163* (0.086)	-0.166* (0.099)
EBT			0.008 (0.166)	-0.099 (0.171)
Kids < 13 X EBT			0.167* (0.091)	0.203** (0.097)
Kids > 12 X EBT			0.203* (0.121)	0.217* (0.132)
Mean of DV	0.80 (1.02)	0.79 (0.97)	0.84 (1.07)	0.79 (1.00)
Policy Switch Only	Y	Y	N	Y
Observations	602	527	1284	1131

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. The amount of the FS transfer, year dummies and family size are included as controls in all specifications. Family size must be explicitly controlled for in these specifications because shifting the unit of observation to the household level removes the fixed-effect.

trolling for the amount of the FS disbursement, family size and adding a year-specific dummy, I then regress number of days fresh vegetables were purchased during a diary on the presence of children, the EBT implementation dummy and the interaction between the two. Results are presented in Table 1.7. While the level of fresh vegetable purchasing is very low (the average household purchases fresh vegetables on about 8% of their shopping days), there is a relationship between children in the household and purchasing that corresponds to the predictions. Prior to EBT implementation, both age classes of children are associated with less frequent purchases of fresh vegetables. Relative to the small levels,

Table 1.8: EBT Erases Heterogeneity Associated with Gender Balance

Food Stamp Type	Coupons	EBT		All	
	(1)	(2)	(3)	(4)	(5)
Days since receipt	-1.480*** (0.362)	-0.750* (0.443)	-1.168*** (0.286)	-1.161*** (0.310)	-0.068*** (0.013)
Days since receipt X Males = Females	-0.857** (0.413)	0.358 (0.382)	-0.632* (0.347)	-0.868** (0.411)	-0.021 (0.015)
Days since receipt X EBT			0.042 (0.321)	-0.093 (0.336)	0.020 (0.017)
Days since receipt X Males = Females X EBT			0.994** (0.495)	1.265** (0.556)	0.019 (0.021)
Policy Switch Only	Y	Y	N	Y	Y
Log Expenses	N	N	N	N	Y
Observations	2415	1881	5000	4296	4296
Clusters	610	533	1296	1141	1141

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. Additionally, I control for household size interacted with both pre- and post-EBT trends such that the estimates refer to the composition rather than the size of the household. Characterization of gender balance is based solely on the adult (age 18 or greater) members of the household. Expenses less than $1 are trimmed to reduce noise, especially important in the log specifications. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results.

the correlation is large. The implementation of EBT has a countervailing effect that almost exactly wipes out the differences in purchasing that come from children.

Moving to the measure of gender balance, the results are similar to those for young children, and consistent with the theory. While the coefficient on the triple interaction in the log specification in column (6) of Table 1.8 is not statistically significant, the point estimate almost fully counteracts the pre-EBT gender-balance interaction. It is worth noting that the magnitude of the un-interacted effect of EBT on the slope of the expenditure profile is non-negligible. One reason for this is that the baseline group includes many families with young children. From column (4), gender-balanced households experience intra-month declines \$0.87 per-day faster than unbalanced households from a baseline of \$1.16 prior to the EBT implementation (or 2.1% on a baseline of 6.8%). The differential is completely counteracted for by the implementation of the EBT program. Results are qualitatively very similar using a variable equal to the absolute value of the difference between the number of male and female adults in the household.

In sum, I find data consistent with the predictions of the preference aggregation model, though not always with statistical confidence. One puzzle is that the household size specification shows little evidence of an interaction with the EBT policy switch or an exacerbating pre-trend. Even though size is not a direct measure of preference variance, increasing it is directly related to whether the model should apply.

1.3.4 Structural Estimation

I perform a calibration of a quasi-hyperbolic model of time preferences in order to assess whether the data generate present-bias estimates in line with the literature, whether the magnitudes of the parametrically-measured heterogeneity in present-bias and EBT effects are meaningful, and then what bargaining weights would need to be in order to generate the observed patterns in the data. First, I make a series of assumptions that allows me to utilize a basic model of intertemporal choice. Consider a household whose budget for food is endowed every 4 weeks and is isolated from all other purchases. The budget is completely exhausted by the end of the 4 week period following receipt and prices do not change. Households derive utility from the log of expenditures on food. The advantage to this setup is that observed expenditures translate directly into value of food consumed and log utility captures fact that consuming no food is extraordinarily undesirable, but marginal utility decreases very quickly such that gourmet food is not worth starving for.

All the simplifying assumptions allow me to use the hyperbolic Euler equation from Harris and Laibson (2001),

$$\frac{c_{t+1}}{c_t} = \beta\delta c'(x_{t+1}) + \delta(1 - c'(x_{t+1})), \tag{7}$$

where consumption levels, c_t are expressed as functions of the current wealth stock, x_t, and β and δ are the present-bias parameter and exponential discount factor from the quasi-hyperbolic model of discounting. Shapiro (2005) proves that this can be reduced to the recursive equation,

$$\alpha_t = \begin{cases} \frac{\alpha_{t+1}}{\alpha_{t+1}+(\delta(1-(1-\beta)\alpha_{t+1}))} & \text{if } t < T \\ 1 & \text{if } t = T \end{cases}, \tag{8}$$

where T represents the last period and α_t is the fraction of wealth spent on consumption in period t.

I borrow the estimated daily $\delta \approx 0.997$ from the Andreoni, List, Savikhin, and Sprenger (2011) data. While this rate is still relatively high, it comes from a context more likely to evaluate individuals' longer-run tradeoff rates rather than their short-run temptation to consume. Since the estimation depends on a recursive definition of consumption and expenditure, I use the predicted values from my fixed-effect regressions to generate complete monthly profiles of expenditures and then collapse by sample period and household characteristics. I include variables that shift the estimated parameter value in the recursive estimation that correspond roughly to the interactions of interest in the main fixed-effect specifications.

Looking at the pooled sample of all households after the introduction of EBT, I estimate $\beta = 0.945$. This is close to other estimates in the literature such as Shapiro (2005), Frederick, Loewenstein, and O'Donoghue (2002), Laibson, Repetto, and Tobacman (2003a, 2003b), Kuhn, Kuhn, and Villeval (2013) and Augenblick, Niederle, and Sprenger (2013). Forcing discounting to be purely exponential generates a yearly discount rate estimate of 51,423,440% (or a daily factor of 0.965). Focusing on single individuls actually produces a β further from 1. Thus, even in the period and population for which preference aggregation should not be a factor, a daily discount rate does not exist.

The estimate of β for a single person household prior to EBT is 0.942. Adding one child, aged 6-12, reduces that estimate all the way to 0.717. This is a substantial effect. After EBT, the single-person estimate is 0.982, or 0.999 with the added 6-12 year old. This gap implies an estimate of $1 - B$, the bargaining weight on the assumed-to-be infinitely impatient child's preferences of about 0.25. For a younger child, the pre-EBT estimate is 0.809 and the post-EBT estimate is 0.977, implying a smaller decision weight of about 0.15. This difference could reflect the fact that very young children cannot bargain as actively. For gender-unbalanced households prior to EBT, the estimate of β is 0.946. Making that household gender-balanced corresponds to reducing β to 0.636. Introducing EBT raises the estimate back up to 0.971.

While these structural estimates should be taken with considerable skepticism due to the many required abstractions and the amount of noise, they underscore two important points. First and foremost, preference aggregation exacerbates the dynamic inconsistency that exists even for single-person households. Second, the exacerbation empirically identified in this paper, regardless of the cause, is economically significant.

1.3.5 Could It be Heterogeneous Stigma?

To the author's knowledge, there is no precise theory or consensus as to how welfare stigma affects families and how it might do so heterogeneously. There are also few empirical facts to draw on. Currie and Grogger (2001) find that SNAP enrollment gains following EBT are limited to rural households and married couples with no children. The lack of gains for families with children suggests that experienced stigma could be based more on deservingness of benefits than parental responsibility. However, stigma from enrolling in SNAP and stigma from using SNAP benefits may not be the same. In the context of benefit usage, common sense dictates that stigma should affect how often benefits are used rather than the value used in any given visit. Thus, policies that reduce the stigma associated with benefit usage could affect shopping frequency in a way that allows for a smoother expenditure profile.

To get directly at this issue, I analyze the frequency of shopping, whether it is affected by EBT and if this effect (or pre-existing trends) indicate heterogeneous stigma.

Table 1.9: Shopping Frequency Unrelated to EBT or Household Composition

hhvar	HH Size (1)	Kids 0-5 Yrs. (2)	Kids 6-12 Yrs. (3)	Kids 13-17 Yrs. (4)	Males = Females (5)
Days since receipt	-0.035***	-0.035***	-0.035***	-0.035***	-0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Days since receipt X EBT	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Days since receipt X *hhvar*	-0.000	0.001	0.000	0.001	0.003
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Days since receipt X *hhvar* X EBT	0.001	0.002	0.003	-0.004	-0.003
	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
Observations	14,809	14,809	14,809	14,809	14,809
Clusters	1302	1302	1302	1302	1302

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effects at the household level and controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. The household size variable is adjusted by subtracting 1, such that the terms non interacted with size apply to single-individual units. Days with expenses less than $1 are considered to be non-shopping days. Household weights change from month to month in the sample, so the specifications presented above use unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results.

Table 1.9 presents regressions of a dummy variable for whether a household made purchases totaling more than $1 on any given day in their diary on the number of days since stamp receipt and other variables of interest.

The uninteracted coefficient on days since receipt is steady across all specifications and none of the interactions are important either statistically or in terms of their magnitude. EBT appears to have no homogeneous or heterogeneous effect on shopping frequency. This analysis casts doubt on the idea that welfare stigma is experienced when benefit usage is observed by other individuals[31] and does not support stigma as an explanation of the EBT

[31]This result has interesting self-signaling versus social-signaling ramifications, but an in-depth analysis is

effects.

1.4 Conclusion

I demonstrate that failures to use food stamps to smooth monthly expenditures on food are more severe for families with young children and gender-balanced homes. The introduction of the EBT program appears to have counterbalanced this discrepancy. I propose that the structure of the theoretical link drawn by Jackson and Yariv (2012a) and Hertzberg (2012) between the literature on household bargaining and time inconsistency is a plausible explanation. The main analysis using expenditures from CEX data is supplemented with supporting evidence on calories consumed from the CSFII on caloric intake. I show that EBT affects bundles purchased in a manner consistent with this hypothesis and try to rule out stigma as an alternative. I view this as a contribution to a new literature at the intersection of behavioral economics and public finance regarding the consequences of program design that are best explained by non-traditional choice theory.

Thaler (1999), writing on mental accounting, describes different types of heuristic wealth, income and consumption accounting that violate the principle of fungibility. The best-known example is the envelope system of budgeting. He notes two purposes for categorical techniques like the envelope system: to facilitate rational tradeoffs between categories and as self-control devices. Thaler also illustrates the downside to the envelope system and similar techniques; when one account is near the limit, individuals will display excess price sensitivity in one category even when there is slack in another account. These

best left for future work.

distortions stem from a violation of basic principles, so it comes as little surprise that they get attention from economists. Indeed, Emily Oster writes in an opinion article for *Slate* (Oster, 2013) that this method of accounting is dangerous as it prevents households from reoptimizing to price changes and clearing bad debt.[32]

While Oster's wisdom applies generally, I argue that the allocation of resources to food in different time periods is an exception to the rule for distressed households. Food is a fundamental unit of daily consumption for which the need fluctuates minimally. The self-control value of the envelopes trumps the fungibility issues, *especially if one person can be put clearly in charge of them.* Treating a food budget as completely non-fungible over a week or two-week horizon could even offer the added benefit of encouraging the consumption of foods that perish on a similar schedule.

A policy implication that follows from this and previous literature on the smoothing of transfer income, is to enforce weekly or biweekly envelopes by shifting disbursement to a higher frequency. Even post-EBT, the uninteracted days since stamps variable is negative, significant and large in magnitude. Beyond that, the results here suggest that property rights play an important role in determining how well transfer income is budgeted over time. The lesson from this finding is broad and echoed by a large development literature on recipient effects. Property rights over other types of transfers should be investigated as well. The Earned Income Tax Credit (EITC) is of particular interest since it is only disbursed only once per year. Whether its use should be smoothed depends on motive: does the EITC break credit constraints for large investments and durable purchases, or does it help bring a

[32]This assumes that households don't reoptimize the size of the envelopes to match price changes. They have self-control value even in a completely static price-setting.

family out of food insecurity over the course of the following year?

A significant caveat to the policy implications is the cost-benefit analysis performed by Shapiro (2005). He accounts for administrative costs and makes assumptions about caloric utility to asses the gains from increasing disbursement frequency and finds little net gain. While I do not disagree with his calculations, I suggest that modeling the impact of improving food stability using a caloric utility function may not be the best approach to quantifying the benefits. Numerous studies, including Hoynes, Page, and Stevens (2011), suggest a positive and significant impact of the Women Infants and Children (WIC) extension of SNAP on nutrient intake and subsequently birthweight even though benefits in 1999 (roughly the middle of my sample period) had an average monthly value of only \$32.50.[33] My estimates indicate that EBT increased the amount spent in a week 4 shopping trip for a family with a young child by about that amount.

Although one of the implications of the preference aggregation theory of dynamic inconsistency is that even time-consistent individuals can become households with present bias, the expenditure data are clear that even households consisting of only one individual display trends that cannot be reconciled with an exponential daily discount rate. The β associated with a quasi-hyperbolic single individual in the CEX sample is 0.927, and the post-EBT β for all households is 0.945. Both estimates represent economically important deviations from 1. Forcing the post-EBT β equal to one would imply a daily exponential discount factor of roughly 0.965, or an annual discount rate of about 51,423,440%. While some households may have a tougher time budgeting because of their structure, in the

[33] http://www.fns.usda.gov/pd/wisummary.htm

population of SNAP recipients, even dictatorial control over resources isn't sufficient to eliminate the calorie crunch.

A final takeaway from this work is that when it comes to programs dedicated to alleviating poverty, assessments made on the yearly or monthly basis may fail to capture important issues. An EITC disbursement in February may bring a household above the poverty line on a yearly-income basis, but one should asses the effectiveness of the program in alleviating poverty based on how many days of the year that household consumes value above the daily poverty threshold.[34] Especially given recent political inclinations to reduce funding for SNAP and other welfare programs, it is worth asking whether behavioral insights into short-run decision making can improve the efficacy of transfers.

[34] Or at least under an assumption of less-than-perfect substitutability of consumption across periods.

1.5 Acknowledgements

Chapter 1, in full, is currently being prepared for submission for publication of the material.

1.6 Appendix

Figure A1.1: EBT Cards Establish the Identity of the Food Stamp Recipient Clearly

New York electronic benefit card (EBT) sample. (New York State)

Figure A1.2: Food Stamp Coupons Were Cash Similar

Obsolete $2.00 Food Coupon. (The National Museum of American History)

Table A1.1: Household Observable Means and Food Stamp Timing

Variable	Week 1	Week 2	Week 3	Week 4	H_0 : All Weeks Equal
Family Size	3.14	3.17	3.14	3.19	$p = 0.932$
Food Stamps ($)	171.53	174.68	175.88	178.74	$p = 0.771$
Adults	1.67	1.64	1.64	1.66	$p = 0.883$
Female Adults	1.12	1.11	1.11	1.11	$p = 0.972$
Male Adults	0.54	0.53	0.53	0.55	$p = 0.899$
Kids 0-5 Yrs.	0.57	0.59	0.60	0.57	$p = 0.883$
Kids 6-12 Yrs.	0.54	0.57	0.56	0.61	$p = 0.536$
Kids 13-17 Yrs.	0.34	0.36	0.33	0.34	$p = 0.849$
FRP Lunches?	0.03	0.03	0.04	0.04	$p = 0.647$

Table A1.2: Expenses on Perishables Decrease throughout the Food-Stamp Month

Dep. Var.	$ / Day		ln($) / Day	$ / Week	ln($) / Week
	(1)	(2)	(3)	(4)	(5)
Periods since receipt	-0.668*** (0.049)	-0.590*** (0.157)	-0.046*** (0.010)	-4.851*** (0.524)	-0.176*** (0.041)
Expenses $\geq$ $1 Only	N	Y	Y	N	Y
Observations	14,809	3399	3399	2918	2004
Clusters	1302	1234	1234	1302	1234

*** $\Rightarrow p < 0.01$, ** $\Rightarrow p < 0.05$, * $\Rightarrow p < 0.10$. Standard errors clustered by household. All models feature fixed-effect at the household level. Columns (1)-(3) have controls for whether any particular day is a weekend, which diary week it comes from and which week of the month it comes from. Household weights change from month to month in the sample, so the specifications presented above are using unweighted data. Using the mean weight or the weight that characterizes the majority of the observations for a household produces very similar results as do OLS or random-effect specifications that utilize the changing weights properly.

Table A1.3: Dates of EBT Completion for States in Sample

Year	States	Eligible for Policy Switch Sample?
1993	Maryland	No
1995	South Carolina, Texas	Yes
1996	Utah	Yes
1997	Alabama, Connecticut, Illinois, Kansas, Louisiana, Massachusetts	Yes
1998	Alaska, Arkansas, Colorado, District of Columbia, Florida, Georgia, Hawaii, Idaho Minnesota, Missouri, Oklahoma, Oregon Pennsylvania, Vermont	Yes
1999	Arizona, Kentucky, New Hampshire, New Jersey, North Carolina, Ohio, Tennessee, Washington	Yes
2000	Wisconsin	Yes
2001	Michigan, New York	Yes
2002	Indiana, Nebraska, Nevada, Virginia	Yes
2003	Iowa	No
2004	California	No

All states eligible for policy switch analysis appear contribute to the analysis except for Arkansas and Idaho.

Table A1.4: Household Observable Means and EBT Timing

Variable	Food Stamp Coupons	EBT Foods Stamps	H_0 : Means Equal
Family Size	3.03	3.19	$p = 0.140$
Food Stamps ($)	170.88	183.98	$p = 0.078$
Adults	1.60	1.67	$p = 0.157$
Female Adults	1.09	1.10	$p = 0.700$
Male Adults	0.51	0.57	$p = 0.136$
Kids 0-5 Yrs.	0.58	0.54	$p = 0.461$
Kids 6-12 Yrs.	0.54	0.59	$p = 0.390$
Kids 13-17 Yrs.	0.31	0.34	$p = 0.397$
FRP Lunches?	0.03	0.03	$p = 0.647$

Sample is limited to states that switched food stamp disbursement methods during the observation period in the CEX data.

Table A1.5: EBT Reallocates Expenditures to Later Weeks Heteregeneously

	Expenditure Difference from Week 2		
	Week 1	Week 3	Week 4
All Households	4.264	-3.751*	-5.561**
	(2.901)	(2.249)	(2.740)
All Households Post EBT	-2.103	3.822	3.674
	(5.210)	(3.524)	(4.738)
Per Kid 0-5 Yrs.	7.796*	-2.177	-4.866
	(4.322)	(2.240)	(3.537)
Per Kid 6-12 Yrs.	5.020	-2.862	-4.206
	(3.863)	(3.451)	(4.396)
Per Kid 13-17 Yrs.	-0.399	1.357	-0.019
	(3.744)	(3.846)	(4.314)
Per Kid 0-5 Yrs. Post EBT	-9.441*	9.907**	10.795*
	(5.702)	(4.548)	(6.048)
Per Kid 6-12 Yrs. Post EBT	-17.546**	14.565***	11.248*
	(8.416)	(5.312)	(6.381)
Per Kid 13-17 Yrs. Post EBT	-4.514	2.796	3.637
	(6.239)	(5.365)	(6.137)

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. Sample consists of 4285 days of expenditures greater than $1 for 1141 households in states that switched disbursement policies in the sample period. Family size interactions with week dummies are controlled for as well as the triple interactions between size, EBT and week dummies. Standard errors clustered by household, household fixed effects included.

Rather than enforce a linear, day-by-day trend over the course of the food stamp month, I can follow Hastings and Washington (2010) and allow for weekly dummies that will give some indication of when heterogeneity is most noticeable. I exclude the second week of the month because the overspending in the first week is of interest. Table A5 presents this specification for the number of children interaction. The effect of EBT is to shift expenditures from week 1 to later weeks in the food stamp month, such that in the post-EBT period families with more young children experience less of a monthly decline.

1.7 References

Mohammed Abdellaoui, Oliver l'Haridon, and Corina Paraschiv. Do couples discount future consequences less than individuals? Working Paper, 2013.

Douglas Almond, Hilary W. Hoynes, and Diane Whitmore Schanzenbach. Inside the war on poverty: The impact of food stamps on birth outcomes. *Review of Economics and Statistics*, 93(2):387–403, 2011.

James Andreoni and Charles Sprenger. Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–3356, 2012.

James Andreoni, Eleanor Brown, and Isaac Rischall. Charitable giving by married couples: Who decides and why does it matter? *Journal of Human Resources*, 38(1):111–133, 2003.

James Andreoni, John A. List, Anya C. Savikhin, and Charles Sprenger. Time preferences in preschool children: Experimental evidence of individual differences. Working Paper, 2011.

James Andreoni, Michael A. Kuhn, and Charles Sprenger. On measuring time preferences. NBER Working Paper No. w19392, 2013.

Sibel Atasoy, Bradford F. Mills, and Christopher F. Parmeter. The dynamics of food stamp program participation: A lagged dependent variable approach. Working Paper, 2010.

Orazio Attanasio and Valerie Lechene. Efficient responses to targeted cash transfers. Forthcoming, *Journal of Political Economy*, 2011.

Orazio Attanasio, Erich Battistin, and Alice Mesnard. Food and cash transfers: Evidence from colombia. *Economic Journal*, 122(559):92–124, 2012.

Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over time: Dynamic inconsistency in real effort tasks. Forthcoming, *Quarterly Journal of Economics*, 2013.

Steven Bednar. Would you like paper or plastic? food stamps disbursement method and take-up. Working Paper, 2011.

B. Douglas Bernheim. Comment on 'family bargaining and retirement behavior'. In Henry Aaron, editor, *Behavioral Economics and Retirement Policy*. Brookings Institution Press, 1999.

Eric Bettinger and Robert Slonim. Patience among children. *Journal of Public Economics*, 91(1-2):343–263, 2007.

Jayanta Bhattacharya and Janet Currie. Youths at nutritional risk: Malnourished or misnourished. In Jonathan Gruber, editor, *Risky Behavior among Youths: An Economic Analysis*. University of Chicago Press, 2000.

Marianne Bitler and Hilary Hoynes. The more things change, the more they stay the same: The safety net, living arrangements, and poverty in the great recession. Working Paper, 2013.

Gustavo J. Bobonis. Is the allocation of resources within the household efficient? new evidence from a randomized experiment. *Journal of Political Economy*, 117(3):453–503, 2009.

Martin Browning and Pierre-André Chiappori. Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, 66(6):1241–1278, 1999.

Martin Browning, François Bourguignon, Pierre-André Chiappori, and Lechene. Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy*, 102(6):1067–1096, 1994.

Fredrik Carlsson, Peter Martinsson, Ping Qin, and Matthias Sutter. The influence of spouses on household decision making under risk: an experiment in rural china. *Experimental Economics*, 16(3):383–401, 2013.

Janet Currie. U.s. food and nutrition programs. In Robert A. Moffitt, editor, *Means-Tested Transfer Programs in the United States*. University of Chicago Press, 2003.

Janet Currie and Jeffrey Grogger. Explaining recent declines in food stamp program participation. In William Gale and Janet Rothenberg-Pack, editors, *Brookings-Wharton Papers on Urban Affairs*. Brookings Institution Press, 2001.

Barbara Devaney and Robert A. Moffitt. Dietary effects of the food stamp program. *American Journal of Agricultural Economics*, 73:202–211, 1991.

Esther Duflo. Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in south africa. *World Bank Economic Review*, 17(1):1–25, 2003.

Shane Frederick, George Loewenstein, and Ted O'Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401, 2002.

Christopher Harris and David Laibson. Dynamic choices of hyperbolic consumers. *Econometrica*, 69(4):935–957, 2001.

Justine Hastings and Ebonya Washington. The first of the month effect: Consumer behavior and store responses. *American Economic Journal: Economic Policy*, 2:142–162, 2010.

Andrew Hertzberg. Exponential individuals, hyperbolic households. Working Paper, 2012.

John Hoddinott and Lawrence Haddad. Does female income share influence household expenditures? evidence from côte d'ivoire. *Oxofrd Bulletin of Economics and Statistics*, 57(1):77–96, 1994.

Hilary Hoynes, Marianne Page, and Ann Huff Stevens. Can targeted transfers improve birth outcomes?: Evidence from the introduction of the wic program. *Journal of Public Economics*, 95(7-8):813–827, 2011.

IHRC. Nourishing change: Fulfilling the right to food in the united states. Technical report, New York University School of Law, 2013.

Matthew O. Jackson and Leeat Yariv. Collective dynamic choice: The necessity of time inconsistency. Working Paper, 2012a.

Matthew O. Jackson and Leeat Yariv. Present bias and collective dynamic choice in the lab. Working Paper, 2012b.

Neeraj Kaushal and Qin Gao. Food stamp program and consumption choices. In Michael Grossman and Naci H. Mocan, editors, *Economic Aspects of Obesity*. University of Chicago Press, 2011.

Michael A. Kuhn, Peter Kuhn, and Marie Claire Villeval. Self control and intertemporal choice: Evidence from glucose and depletion interventions. Working Paper, 2013.

David Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–477, 1997.

David Laibson, Andrea Repetto, and Jeremy Tobacman. A debt puzzle. In Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, editors, *Knowledge, Information and Expectations in Modern Economics: In Honor of Edmund S. Phelps*. Princeton University Press, 2003a.

David Laibson, Andrea Repetto, and Jeremy Tobacman. Wealth accumulation, credit card borrowing, and consuption-income comovement. Working Paper, 2003b.

Steven D. Levit, John A. List, Susanne Neckermann, and Sally Sadoff. The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. NBER Working Paper No. 18165, 2012.

Shelly J. Lundberg, Robert A. Pollak, and Terence J. Wales. Do husbands and wives pool their resources? evidence from the united kingdom child benefit. *Journal of Human Resources*, 32(3):463–480, 1997.

Giovanni Mastrobuoni and Matthew Weinberg. Heterogeneity in intra-monthly consumption patterns, self-control, and savings at retirement. *American Economic Journal: Economic Policy*, 1:163–189, 2009.

W. Mischel, E.B. Ebbesen, and A. Raskoff Zeiss. Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, 21(2):204–218, 1972.

Ted O'Donoghue and Matthew Rabin. Doing it now or later. *American Economic Review*, 89(1):103–124, 1999.

Emily Oster. Household deficit: Youre probably doing your household budget all wrong. *Slate*, 2013.

Jesse Shapiro. Is there a daily discount rate? evidence from the food stamp nutrition cycle. *Journal of Public Economics*, 89:303–325, 2005.

Melvin Stephens Jr. '3rd of tha month': Do social security recipients smooth consumption between checks? *American Economic Review*, 93:406–422, 2003.

Melvin Stephens Jr. Paycheque receipt and the timing of consumption. *Economic Journal*, 116:680–701, 2006.

Richard H. Thaler. Mental accounting matters. *Journal of Bheavioral Decision Making*, 12(3):183–206, 1999.

Shing-Yi Wang. Property rights and intra-household bargaining. Working Paper, 2012.

Stéphane Zuber. The aggregation of preferences: Can we ignore the past? *Theory and Decision*, 70(3):367–384, 2011.

CHAPTER 2

On Measuring Time Preferences[35]

Abstract:

Eliciting time preferences has become an important component of both laboratory and field experiments, yet there is no consensus as how to best measure discounting. We examine the predictive validity of two recent, simple, easily administered, and individually successful elicitation tools: Convex Time Budgets (CTB) and Double Multiple Price Lists (DMPL). Using similar methods, the CTB and DMPL are compared using within- and out-of-sample predictions. While each perform equally well within sample, the CTB significantly outperforms the DMPL on out-of-sample measures.

[35]Co-authored with James Andreoni, of the University of California, San Diego, and Charles Sprenger of Stanford University.

2.1 Introduction

Time preferences are fundamental to theoretical and applied studies of decision-making, and are a critical element of much of economic analysis. At both the aggregate and individual level, accurate measures of discounting parameters can provide helpful guidance on the potential impacts of policy and provide useful diagnostics for effective policy targeting.

Though efforts have been made to identify time preferences from naturally occurring field data,[36] the majority of research has relied on laboratory samples using variation in monetary payments.[37] Despite many attempts, however, the experimental community lacks a clear consensus on how best to measure time preferences; a point made clear by Frederick, Loewenstein and O'Donoghue (2002). One natural challenge which has gained recent attention is the confounding effect of utility function curvature. Typically, linear utility is assumed for identification, invoking expected utility's necessity of linearity for small stakes decisions (Rabin, 2000). In an important recent contribution, Andersen et al.

[36]These methods investigate time preferences by examining durable goods purchases, consumption profiles or annuity choices (Hausman, 1979; Lawrance, 1991; Warner and Pleeter, 2001; Gourinchas and Parker, 2002; Cagetti, 2003; Laibson, Repetto and Tobacman, 2003, 2007). While there is clear value to these methods they may not be practical for field settings with limited data sources or where subjects make few comparable choices.

[37]Chabris, Laibson and Schuldt (2008) identify several important issues related to this research agenda, calling into question the mapping from experimental choice to corresponding model parameters in monetary discounting experiments. Paramount among these issues are clear arbitrage arguments such that responses in monetary experiments should reveal only the interval of borrowing and lending rates, and thus limited heterogeneity in behavior if subjects face similar credit markets (Cubitt and Read, 2007; Andreoni and Sprenger, 2012a, 2012b). This last concern may be beyond the reach of most experimental samples. Evidence from Coller and Williams (1999) suggests that even when the entire arbitrage argument is explained to subjects, heterogeneity remains and responses do not collapse to reasonable intervals of borrowing and lending rates. Following most of the literature, the experiments we conduct will focus on monetary choices, taking the laboratory offered rates as the relevant ones for choice. Importantly, the methods we describe are easily portable to other domains with less prominent fungibility problems. One recent example using the Convex Time Budget described below with choices over effort is Augenblick, Niederle, and Sprenger (2013).

(2008) show that if utility is assumed to be linear in experimental payoffs when it is truly concave, estimated discount rates will be biased upwards.[38] This observation has reset the investigation of new elicitation tools.

Andersen, et al. (2008) (henceforth AHLR) offer the clever use of measures of risk taking to incorporate utility function curvature, which we refer to as a Double Multiple Price List (DMPL). Andreoni and Sprenger (2012a) (henceforth AS) used variation in linear budget constraints over early and later consumption to identify convexity of preferences, a device they call a Convex Time Budget (CTB). Both methods show promise, as each is easy to implement in both the lab and the field, and both are successful at tightly estimating "credible" values of key parameters. The objective of this study is to work toward a consensus by comparing these two methods.

One criterion should, obviously, be simplicity. In particular, researchers eliciting preferences put a premium on devices that are simple for subjects, easy to administer, transportable to the field, and can be easily folded into a larger research design. Both methods seem to succeed equally well on this dimension.

More central to our analysis, we propose that predictive validity as the the second and most relevant criterion. In particular, parameter estimates generated from a specific data set should yield good in-sample fit, have out-of-sample predictive power, and predict

[38]Frederick, Loewenstein and O'Donoghue (2002) also provide discussion of this confound and present three strategies for disentangling utility function curvature from time discounting: 1) eliciting utility judgments such as attractiveness ratings at two points in time; 2) eliciting preferences over temporally separated probabilistic prospects to exploit the linearity-in-probability property of expected utility; and 3) "separately elicit the utility function for the good in question, and then use that function transform outcome amounts to utility amounts, from which utility discount rates could be compute" (p. 382). The third of these techniques is close in spirit to the Double Multiple Price List implemented by Andersen et al. (2008) described below.

relevant, genuine economic activity.[39]

We document two main findings when examining predictive validity. First, we reproduce the broad conclusions of both AHLR and AS, that is, there are clear confounding effects of curvature that need to be controlled for in estimating discounting. Second, when taking these estimates out-of-sample we find that the CTB-based estimates markedly outperform the DMPL-based estimates when predicting intertemporal choice. We show that the key driver of these results is the different assumptions employed to identify utility function curvature–risk aversion (DMPL) versus neo-classical demand theory (CTB). Interestingly, we find that the relatively large number of corner solutions observed with the CTB (an aspect of the method criticized by Harrison, Lau and Rutström, 2013) is precisely the quality of the CTB that generates the greatest improvement in predictive power over the DMPL. The DMPL greatly over-predicts interior solutions in the CTB, while the CTB is equally good as the DMPL at predicting choices on the MPL for time allocations. In predicting other experimental choices over time, the DMPL predictions offer no marginal explanatory power, while CTB estimates are significant predictors.

Section 2.2 describes our preference elicitation techniques and experimental protocol. Section 2.3 presents estimation results and evaluates the success of the CTB and DMPL at predicting choice both in- and out-of-sample. Section 2.4 concludes.

[39]Though this seems a natural objective, there are relatively few examples of research linking laboratory measures of time preference to other behaviors or characteristics (Mischel, Shoda and Rodriguez, 1989; Ashraf, Karlan and Yin, 2006; Dohmen et al., 2010; Meier and Sprenger, 2010, 2012). These exercises at times demonstrate the lack of explanatory power for prior time preference estimates (Chabris et al., 2008).

2.2 Techniques and Protocol

Before introducing the two considered elicitation techniques, we first outline the nature of preferences. Consider allocations of experimental payments, x_t and x_{t+k} between two periods, t and $t + k$. Preferences over these experimental payments are assumed to be captured by a stationary, time-independent constant relative risk averse utility function $u(x_t) = x_t^\alpha$. We assume a quasi-hyperbolic structure for discounting (Laibson, 1997; O'Donoghue and Rabin, 1999), such that preferences over bundles are described by

$$U(x_t, x_{t+k}) = \begin{cases} x_t^\alpha + \beta\delta^k x_{t+k}^\alpha & \text{if } t = 0 \\ x_t^\alpha + \delta^k x_{t+k}^\alpha & \text{if } t > 0. \end{cases} \tag{1}$$

The parameter δ captures standard long-run exponential discounting, while the parameter β captures a specific preference towards payments in the present, $t = 0$. The one period discount factor between the present and a future period is $\beta\delta$, while the one period discount factor between two future periods is δ. Present bias is associated with $\beta < 1$ and $\beta = 1$ corresponds to the case of standard exponential discounting.[40]

We consider two elicitation techniques, the DMPL and the CTB, designed to provide identification of the three parameters of interest, α, δ, and β, corresponding to utility function curvature, long-run discounting, and present bias, respectively. Given that any functional form of utility one estimates will be misspecified to some degree, different methods are likely to yield different parameter estimates. While these differences are important,

[40]We abstract away from any discussion of sophistication or naiveté wherein individuals are potentially aware of their predilection of being more impatient in the present than they are in the future. Our implemented experimental techniques will be unable to distinguish between the two.

our view us that the first concern is to have a method that is useful as a predictive tool for the researcher community.

2.2.1 Elicitation Techniques

We begin by presenting the DMPL, which consists of two stages. The first stage is designed to identify discounting, potentially confounded by utility function curvature. The second stage is designed to un-confound the first stage by providing information on utility function curvature through decisions on risky choice. In the first stage, individuals make a series of binary choices between smaller sooner payments and larger later payments. Such binary choices are organized into Multiple Price Lists (MPL) in order of increasing gross interest rate (Coller and Williams, 1999; Harrison Lau and Williams, 2002). The point in each price list where an individual switches from preferring the smaller sooner payment to the larger later payment carries interval information on discounting. Figure 2.1, Panel A, presents a sample intertemporal MPL.[41]

Importantly, one cannot make un-confounded inference for time preferences based on these intertemporal responses alone. Consider an individual who prefers \$X at time t over \$Y at time $t+k$, but prefers \$Y at time $t+k$ over \$X'<\$X at time t. If $t \neq 0$ then one can infer the bounds on δ to be $\delta \in (X'^{\alpha}/Y^{\alpha}, X^{\alpha}/Y^{\alpha})$. Though standard practice for identifying δ often (at times implicitly) assumes linear utility, $\alpha = 1$, it's clear that a concave utility function, $\alpha < 1$, will bias discount factor estimates downwards,

[41] This implementation appears slightly different from others for coherence with our implementation of the CTB. In effect, individuals choose between smaller sooner payments and larger later payments. However, we clarify that choosing the smaller sooner payment implies a subject will receive zero at the later date, and vice versa.

TODAY *and* 5 WEEKS from today

For each decision number (1 to 6) below, decide the **AMOUNTS** you would like for sure **today** **AND** **in 5 weeks** by checking the corresponding box.
Example: In Decision 1, if you wanted $19.00 today and $0 in five weeks you would check the left-most box. Remember to check only one box per decision!

1.	payment TODAY … *and* payment in 5 WEEKS	$19.00 $0 ☐	$0 $20.00 ☐
2.	payment TODAY … *and* payment in 5 WEEKS	$18.00 $0 ☐	$0 $20.00 ☐
3.	payment TODAY … *and* payment in 5 WEEKS	$17.00 $0 ☐	$0 $20.00 ☐
4.	payment TODAY … *and* payment in 5 WEEKS	$16.00 $0 ☐	$0 $20.00 ☐
5.	payment TODAY … *and* payment in 5 WEEKS	$14.00 $0 ☐	$0 $20.00 ☐
6.	payment TODAY … *and* payment in 5 WEEKS	$11.00 $0 ☐	$0 $20.00 ☐

Panel A: Intertemporal Multiple Price List

Decision	Option A						Option B					
		If the die reads	you receive	and	If the die reads	you receive		If the die reads	you receive	and	If the die reads	you receive
1	☐	1	15		2-10	8.31	☐	1	20		2-10	0.52
2	☐	1-2	15		3-10	8.31	☐	1-2	20		3-10	0.52
3	☐	1-3	15		4-10	8.31	☐	1-3	20		4-10	0.52
4	☐	1-4	15		5-10	8.31	☐	1-4	20		5-10	0.52
5	☐	1-5	15		6-10	8.31	☐	1-5	20		6-10	0.52
6	☐	1-6	15		7-10	8.31	☐	1-6	20		7-10	0.52
7	☐	1-7	15		8-10	8.31	☐	1-7	20		8-10	0.52
8	☐	1-8	15		9-10	8.31	☐	1-8	20		9-10	0.52
9	☐	1-9	15		10	8.31	☐	1-9	20		10	0.52
10	☐	1-10	15		-	8.31	☐	1-10	20		-	0.52

Panel B: Holt-Laury Risk Elicitation

Figure 2.1: Sample DMPL Decision Sheets

understating the true bounds.[42] Further, without some notion of the extent of curvature, one cannot un-confound the measure. This motivates the second stage.

The second stage of the DMPL is designed to account for utility function curvature by introducing a second experimental measure. In particular, a Holt and Laury (2002, henceforth HL) risk preference task is conducted alongside the intertemporal decisions. Subjects face a series of decisions between a safe and a risky binary gamble. The probability of the high outcome in each gamble increases as one proceeds through the task, such that where a subject switches from the safe to the risky gamble carries information on risk attitudes. Figure 2.1, Panel B, presents a sample HL task. The risk attitude elicited in the HL task identifies the degree of utility function curvature, α, which is then applied to the intertemporal choices to un-confound the discounting bounds. In effect, α is identified from risky choice data, and δ and β are identified from intertemporal choice data.

The CTB takes a different approach to identification. Instead of incorporating a second experimental elicitation, the CTB recognizes a key restriction of the standard multiple price list approach. When making a binary choice between a smaller sooner payment, \$X, and a larger later payment, \$Y, subjects are effectively restricted to the corner solutions in (sooner, later) space, (\$X, \$0) and (\$0, \$Y). That is, they maximize the utility function in (1) subject to the discrete budget constraint $(x_t, x_{t+k}) \in \{(X, 0), (0, Y)\}$. If the utility function is indeed linear, such that $\alpha = 1$, the restriction to corners is non-binding.[43]

[42]Correspondingly, a convex utility function biases discount factors upwards. A similar issues exists for identifying β when $t = 0$.

[43]A key caveat to this is that while the restriction of the data to corner solutions in non-binding in the case of linear utility, it does not mean that the same set of choices on restricted and unrestricted data will yield the same parameter estimates. Corner choices on unrestricted data have very different implications for preferences than corner choices on restricted data. This turns out to be key for our results.

TODAY *and* 5 WEEKS from today

For each decision number (1 to 6) below, decide the **AMOUNTS** you would like for sure **today** **AND** **in 5 weeks** by checking the corresponding box.
Example: In Decision 1, if you wanted $19.00 today and $0 in five weeks you would check the left-most box. Remember to check only one box per decision!

1.	payment TODAY …	$19.00	$15.20	$11.40	$7.60	$3.80	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐
2.	payment TODAY …	$18.00	$14.40	$10.80	$7.20	$3.60	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐
3.	payment TODAY …	$17.00	$13.60	$10.20	$6.80	$3.40	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐
4.	payment TODAY …	$16.00	$12.80	$9.60	$6.40	$3.20	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐
5.	payment TODAY …	$14.00	$11.20	$8.40	$5.60	$2.80	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐
6.	payment TODAY …	$11.00	$8.80	$6.60	$4.40	$2.20	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐

Figure 2.2: Sample CTB Decision Sheet

However, if $\alpha < 1$, individuals have convex preferences in (sooner, later) space, preferring interior solutions, and leading the restriction to corners to meaningfully restrict behavior.

This observation leads to a natural solution. If one wishes to identify preferences in (sooner, later) space, one can convexify the decision environment. In a CTB, subjects are given the choice of ($X, $0), ($0, $Y) or anywhere along the intertemporal budget constraint connecting these points such that $Px_t + x_{t+k} = Y$, where $P = \frac{Y}{X}$ represents the gross interest rate. Figure 2.2 presents a sample CTB allowing for interior solutions between the two corners.[44] In the CTB, sensitivity to changing interest rates delivers identification of α

[44] Notably, the version of the CTB we use is different than that of AS. AS used a computer interface to offer individuals 100 tokens that could be allocated to the sooner or later payoffs in any proportion. By condensing the budget to 6 options, we can represent the choice in a check-the-box format that fits onto a sheet of paper. While information is lost in this discretization, it puts the CTB on the same footing as the DMPL in terms of ease-of-administration and portability.

while variation in the timing of payments identifies the discounting parameters, β and δ.[45]

The most important distinction between the two methods is the source of identification of curvature. The DMPL identifies utility function curvature based on the degree of risk aversion elicited in the HL risky choice. In contrast, the CTB identifies curvature based on the degree of price sensitivity in intertemporal choice. These varying sources of information for the shape of the utility function should be equivalent under the utility formulation in (1). The parameter α determines both the extent of intertemporal substitution and the extent of risk aversion.[46] However, there may be reason to expect differences in the extent of measured utility function curvature and hence discounting estimates across the two methods. AHLR document substantial utility function curvature in HL tasks, leading to substantial changes in discounting estimates when accounted for in the DMPL. In contrast, AS document substantially less utility function curvature from CTB choices.[47]

2.2.2 Experimental Design

In order to assess the predictive validity of the DMPL and CTB elicitation methods, we designed a simple within-subject experiment. Subjects faced 4 intertemporal MPLs, 2 HL risk tasks, and 4 CTBs of the form presented in Figures 2.1 and 2.2. For the intertemporal decisions the CTBs and MPLs took the exact same start dates, t, delay lengths, k, and gross interest rates, P. The experimental budget was always \$20 such that the intertem-

[45]This is shown explicitly in section 2.2.3.

[46]Provided α is the sole source of curvature and expected utility maintains in atemporal choice.

[47]However, the AS estimates do differ significantly from linear utility. Further AS show that the extent of CTB utility function curvature is correlated with the distance between standard price list discount factor estimates and CTB discount factor estimates. Individuals with more concave CTB-measured utility functions have more downwards-biased discount factor price list estimates.

Table 2.1: Intertemporal Experimental Parameters

Choice Set	t (initial delay)	k (delay)	P (price ratios): $Px_t + x_{t+k} = 20$
CTB_1, MPL_1	0	35	1.05, 1.11, 1.18, 1.25, 1.43, 1.82
CTB_2, MPL_2	0	63	1.00, 1.05, 1.18, 1.33, 1.67, 2.22
CTB_3, MPL_3	35	35	1.05, 1.11, 1.18, 1.25, 1.43, 1.82
CTB_4, MPL_4	35	63	1.00, 1.05, 1.18, 1.33, 1.67, 2.22

The price ratios for $k = 35$ correspond to yearly (compounded quarterly) interest rates of 65%, 164%, 312%, 529%, 1301% and 4276%. The price ratios for $k = 63$ correspond to rates of 0%, 33%, 133%, 304%, 823% and 2093%.

poral budget constraint in each decision was $Px_t + x_{t+k} = 20$. Hence, as presented in Figures 2.1 and 2.2, the only difference between the implemented CTBs and MPLs was the presence of interior allocations. Table 2.1 summarizes the parameters of the intertemporal choice portion of the experiment. The interest rates, experimental budgets and delay lengths are chosen to be comparable to those of AS. As presented in Figure 2.1, Panel B, in the two HL tasks subjects faced a series of decisions between a safe and a risky gamble. In the first HL task, HL_1, the safe gamble outcomes were \$10.39 and \$8.31, while the risky gamble outcomes were \$20 and \$0.52. In the second HL task, HL_2, the safe gamble outcomes were \$13.89 and \$5.56, while the risky gamble outcomes were \$25 and \$0.28. These values were chosen to provide a measure of curvature at monetary payment values close to those implemented in the intertemporal choices and are scaled versions of those used in the original HL tasks.[48]

Our sample consists of 64 undergraduates, evenly divided into 4 sessions, conducted in February of 2009. Upon arriving in the laboratory, subjects were told they would

[48] See appendix section 2.5.7 for the full instructions. In the HL baseline task, the safe gamble outcomes were \$2.00 and \$1.60 and the risky gamble outcomes were \$3.85 and \$0.10. Our HL_1 scales the largest payment to \$20 and keeps all ratios the same. The second task, HL_2, increases the highest payment to \$25 and increases the variance.

be participating in an experiment about decision-making over time. Subjects were told that based on the decisions they made, and chance, they could receive payment as early as the day of the experiment, as late as 14 weeks from the experiment, or other dates in between. All of the payments dates were selected to avoid holidays or school breaks, and all payments were designed to arrive on the same day of the week. All choices were made with paper and pencil and the order in which subjects completed the tasks was randomized. Two orders were implemented with the HL tasks acting as a buffer between the more similar time discounting choices: 1) MPL, HL, CTB; 2) CTB, HL, MPL.[49] Subjects were told that in total they would make 49 decisions. One of these decisions would be chosen as the 'decision-that-counts' and their choice would be implemented.[50] The full instructions are provided in appendix section 2.5.7.

A primary concern in the design of discounting experiments is to equalize all transaction costs between different dates of payment. Eliminating any uncertainty over delayed payments and convenience of immediate payments is key to obtaining accurate results. We follow the techniques used in AS and take six specific measures to equate transaction costs and ensure payment reliability.[51] Subjects were surveyed extensively after the completion

[49]No order effects were observed.

[50]Our randomization device for implementing the decision-that-counts favored the intertemporal choices over the HL choices. Whereas each time preference allocation was viewed as a choice (48 in total), the HL tasks were viewed as a single choice. When the HL tasks were explained, subjects were told that if these were chosen as the decision-that-counts, then a specific HL choice would be picked at random (with equal likelihood) and a 10-sided dice would be rolled to determine lottery outcomes. Payment would be made in cash immediately in the lab, and subjects would receive a show-up fee of $10 immediately as well. We recognize that this favored randomization may limit the attention subjects pay to the HL tasks. Our results, however, are comparable to other findings of risk aversion in Holt and Laury (2002) and to other implementations of the DMPL (Andreoni and Sprenger, 2012b).

[51]As in AS, all participants lived on campus at UC San Diego, which meant that they had 24 hour access to a locked personal mailbox. Our first measure was to use these mailboxes for intertemporal payments. Second, intertemporal payments were made by personal check from Professor James Andreoni. Although this introduces a transaction cost, it ensures an equal cost in all potential periods of distribution. In addition,

of the experiment. Importantly, 100% of subjects said that they believed that their earnings would be paid out on the appropriate dates.

Once the decision-that-counts was chosen, subjects participated in a Becker, Degroot and Marschak (1964, henceforth BDM) auction eliciting their lowest willingness to accept in their sooner payment to forgo a claim to an additional \$25 in their later payment with a uniform distribution of random prices drawn from $[\$15.00, \$24.99]$.[52] The instructions outlined the procedure and explicitly informed subjects that "the best idea is to write down your true value...".[53] Subsequently, subjects completed a survey including demographic details as well as two hypothetical measures of patience. The first hypothetical measure asked subjects to state the dollar amount of money today that would make them indifferent to \$20 in one month. The second hypothetical measure asked subjects to state the mount of money in one month that would make them indifferent to \$20 today.[54]

While there were 64 subjects in total, our estimation sample for the remainder of the paper consists of 58 individuals. Five individuals exhibited multiple switching at some point in the HL task. One individual never altered their decision from a specific corner solution in all 4 CTBs and thus provided insufficient variation for the calculation of utility

these checks were drawn on an account at the on-campus credit union. Third, for intertemporal payments the \$10 show-up fee was split into two \$5 minimum payments avoiding subjects loading on one experimental payment date to avoid cashing multiple checks. Fourth, the payment envelopes were self-addressed, reducing risk of clerical error. Fifth, subjects noted payment amounts and dates from the decision-that-counts on their payment envelopes, eliminating the need to recall payment values and reducing the risk of mistaken payment. Sixth, all subjects received a business card with telephone and e-mail contacts they could use in case a payment did not arrive. Subjects were made aware of all of these measures prior to the choice tasks.

[52] Subjects were potentially aware of their payment amounts at this point if they remembered their choice exactly.

[53] This follows the protocol of Ariely, Loewenstein and Prelec (2003). A copy of the elicitation and instructions can be found in appendix section 2.5.7.

[54] The exact wording of the first question was 'What amount of money, \$X, if paid to you today would make you indifferent to \$20 paid to you in one month?' The exact wording of the first question was 'What amount of money, \$Y, would make you indifferent between \$20 today and \$Y one month from now?'

parameters. These 6 subjects are dropped to maintain a consistent number of observations across estimates.

2.2.3 Parameter Estimation Strategies

The data collected in the experiment are used to separately identify the key parameters of utility function curvature, α, discounting, δ, and present bias, β for both the CTB and the DMPL. Preferred estimation strategies for recovering these parameters differ between the two elicitation techniques. The CTB is akin to maximizing discounted utility subject to a future value budget constraint. Hence, a standard intertemporal Euler equation maintains,

$$MRS = \frac{x_t^{\alpha-1}}{\beta^{t_0}\delta^k x_{t+k}^{\alpha-1}} = P,$$

where t_0 is an indicator for whether $t = 0$. This can be rearranged to be linear in our experimental variations, t, k, and P,

$$ln\left(\frac{x_t}{x_{t+k}}\right) = \frac{ln(\beta)}{\alpha-1}t_0 + \frac{ln(\delta)}{\alpha-1}k + \frac{1}{\alpha-1}ln(P). \tag{2}$$

Assuming an additive error structure, this is estimable at either the group or individual level, with parameters of interest recovered via non-linear combinations of regression coefficients and standard errors calculated via the delta method. Equation (2) makes clear the mapping from the variation of experimental parameters to structural parameter estimates. Variation in the gross interest rate, P, delivers the utility function curvature, α. For a fixed interest rate, variation in delay length, k, delivers δ, and variation in whether the present, $t = 0$, is considered delivers β.

Three natural issues arise with the estimation strategy described above. First, the allocation ratio $ln\left(\frac{x_t}{x_{t+k}}\right)$ is not well defined at corner solutions.[55] Second, even if the optimality condition were defined at corner solutions, the preferences we assume cannot generate such choices in the form of point-identified maxima. Indeed, this issue is a common point of criticism of CTB approaches (Harrison et al. 2013). Third, this strategy effectively ignores the interval nature of the data, created by the discretization of the budget constraint.

To address the first issue, one can use the demand function to generate a non-linear regression equation based upon

$$x_t = \frac{20(\beta^{t_0}\delta^k P)^{\frac{1}{\alpha-1}}}{1 + P(\beta^{t_0}\delta^k P)^{\frac{1}{\alpha-1}}}, \tag{3}$$

which avoids the problem of the logarithmic transformation in (2). However, this demand function is only defined for $\alpha \in (0,1)$, so the use of either of these techniques is still subject to bias incurred by the second issue above.[56] While this is issue is minimized by the fact that our metric for success is predictive validity, we propose a third technique, Interval Censored Tobit (ICT) regression, that is robust to all three issues mentioned above. While this technique is less transparent and more complicated to perform, it serves as a robustness check for approaches (2) and (3). The details are discussed in appendix section 2.5.1.[57]

[55] In our application we solve this issue operationally, by transforming the \$0 payment in a corner solution to \$0.01 such that the log allocation ratio is always well-defined. Additionally, we consider exercises adding in the fixed \$5 minimum payments to each payment date and qualitatively similar results. See Table A2.2 in the appendix.

[56] Assuming that the degree of misspecification depends on the experimentally varied parameters to some degree, this will be problematic.

[57] AS provide a variety of estimates using both demand functions and Euler equations and several utility formulations such as CARA and CRRA. Broadly consistent estimates are found across techniques.

Preferred methodology for estimating intertemporal preference parameters from DMPL data, as per AHLR, relies on maximum likelihood methods. Binary choices between \$X sooner and \$Y later are assumed to be guided by the utilities $U_X = \delta^t X^\alpha$ and $U_Y = \beta^{t_0} \delta^{t+k} Y^\alpha$. AHLR assign choice probabilities using Luce's (1959) formulation based on these utility values

$$Pr(Choice = X) = \frac{U_X^{\frac{1}{\nu}}}{U_X^{\frac{1}{\nu}} + U_Y^{\frac{1}{\nu}}}, \tag{4}$$

where ν represents stochastic decision error. As ν tends to infinity all decisions become random and as ν tends to zero, all decisions are deterministic based on the assigned utilities. The log of this choice probability represents the likelihood contribution of a given observation.

In order to simultaneously estimate utility function curvature and discounting parameters, AHLR also define a similar likelihood contribution for a HL risk task observation, constructed under expected utility. An alternate stochastic decision error parameter, μ, is estimated for risky choice. As in AHLR, we provide estimates based on only the intertemporal decisions, assuming $\alpha = 1$, and on the combination of time and risk choices. We additionally provide estimates using only the risky data to demonstrate the extent to which estimated utility function curvature is informed by the HL choices. Appendix section 2.5.2 provides full detail of the maximum likelihood strategies for DMPL data.

A subtle, but critical difference between these estimation strategies is how choice 'errors', instances in which the option with the highest utility conditional on the estimated parameters is not selected, occur. Errors enter the CTB specification nested in the context of optimality: unobserved mean-zero shocks specific to one decision that perturb the tangency condition from what would be expected based on estimated parameters. In the DMPL framework, 'errors' come from estimated parameters, ν and μ, that are constant across the estimation sample, and represent how deterministic the relationship is between utility, conditional on estimated parameters, and choice. An econometric model of probabilistic choice cannot be derived from a model of economic optimization without the use of a specialized distributional assumption on the unobservables. If one is concerned about the applicability of the estimates to a more general choice space, it is worth carefully evaluating the preferred source of the structural assumptions that provide identification. We return to this issue in Section 2.3.2.3.

2.3 Results

We present the results in two stages. First, we provide estimation results based on the DMPL and CTB elicitation techniques, drawing some contrasts between the parameter estimates across the two methods. Second, we move to choice prediction and conduct two complementary analyses, attempting to predict choice across methods and attempting to predict choice out-of-sample to our BDM and hypothetical choice data.

2.3.1 Parameter Estimates

Our main estimation results are presented in Table 2.2, providing aggregate estimates of α, β, and an annualized discount rate $r = \delta^{-365} - 1$ for both elicitation techniques and the variety of estimation strategies described in section 2.2.3.[58] Standard errors are clustered on the individual level. To begin, in columns (1) and (2) we separately analyze the two components of the DMPL. In column (1), we assume linear utility and use the intertemporal choice data to estimate β and r. When assuming linear utility, we estimate an annual discount rate of 102.2 percent (s.e. 22.3 percent). In column (2), we use only the HL data to estimate utility function curvature, estimating α of 0.549 (0.044), comparable to other experimental findings on the extent of small stakes risk aversion (e.g., Holt and Laury, 2002). Based on this curvature estimate, an individual would be indifferent between a 50-50 gamble over \$20 and \$0 and \$5.67 for sure, implying a risk premium of \$4.33. The extent of concavity found in column (2) suggests that the estimated annual discount rate of 102 percent in column (1) is dramatically upwards-biased. In column (3) we use both elements of the DMPL to simultaneously estimate utility function curvature and discounting. Indeed, we find that the estimated annual discount rate falls dramatically to 47.2 percent (10.3 percent). The difference in discounting with and without accounting for curvature is significant at all conventional levels, ($\chi^2(1) = 15.71,\ p < 0.01$). This finding echoes those of AHLR, though our estimated discount rates are higher in general. Note that the curvature estimate is virtually identical across columns (2) and (3), indicating the extent to which the measure is informed by risky choice responses.

Next, we consider the CTB estimates. Table 2.2, columns (4) - (6) contain estimates based on the three methods described in section 2.2.3. In column (4), ordinary least squares

[58] We also estimate the parameters of interest on an individual level. Median estimates correspond generally to those in Table 2.2. These results and additional discussion are found in appendix section 2.5.3

Table 2.2: Aggregate Utility Parameter Estimates

	Discounting	Curvature	Discounting and Curvature			
Elicitation:	MPL	HL	DMPL		CTB	
Estimation:	ML	ML	ML	OLS	NLS	ICT
	(1)	(2)	(3)	(4)	(5)	(6)
Utility Parameters						
r	1.022		0.472	0.741	0.679	0.630
	(0.223)		(0.103)	(0.390)	(0.148)	(0.230)
β	0.986		0.992	1.010	0.988	0.997
	(0.010)		(0.006)	(0.022)	(0.009)	(0.016)
α		0.549	0.549	0.947	0.928	0.867†
		(0.044)	(0.044)	(0.003)	(0.007)	(0.017)
Error Parameters						
ν	0.085		0.046			
	(0.010)		(0.007)			
μ		0.096	0.096			
		(0.010)	(0.010)			
Clusters	58	58	58	58	58	58
N	1392	1160	2552	1392	1392	1392
Log Likelihood	-546	-327	-873			-2102
R^2				0.401	0.591	

†: The ICT estimate for α is only identified up to a constant. See appendix section 2.5.1 for details. Standard errors clustered at the individual level in parentheses. Each individual made 20 decisions on the HL, 24 decision on the MPL (and therefore 44 decisions on the DMPL) and 24 decisions on the CTB. In columns (1) through (3) HL, MPL and DMPL estimates are obtained via maximum likelihood using Luce's (1959) stochastic error probabilistic choice model. The CTB is estimated in three different ways: ordinary least squares (OLS) using the Euler equation (2), non-linear least squares (NLS) using the demand function (3) and interval-censored tobit (ICT) maximum likelihood using the Euler equation (2). All maximum likelihood models are estimated using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm.

estimates based on the Euler equation (2) are presented.[59] The annual discount rate is estimated to be 74.1 percent (s.e. 39 percent), generating wide intervals for the extent of

[59]The dependent variable is taken to be the chosen option in all interior allocations. For corner solutions in order for the log allocation ratio to be well defined we transform the value \$0 to \$0.01.

discounting. Hence, the discounting estimate from the DMPL method would lie in the 95 percent confidence interval of the CTB estimate. Importantly, the estimates of utility function curvature in column (4) are far closer to linear utility than that obtained from the DMPL. Based on CTB methods, we estimate α of 0.947 (0.003). With this level of curvature, an individual would be indifferent between a 50-50 gamble over \$20 and \$0 and \$9.62 for sure, implying a risk premium of \$0.38. Column (5) provides non-linear least squares estimates based on the demand function (3). Broadly similar findings are obtained. Column (6) presents interval censored tobit estimates based on the Euler equation (2), accounting for the interval nature of the response data. We draw attention to the estimate of α, which is not directly comparable to our other estimates as this parameter is only identified up to a constant of proportionality (see appendix section 2.5.1 for detail). Beyond this difference, similar estimates for discounting parameters are obtained. Though our estimated discount rates are higher than those of AS, broad consistency in discounting and curvature estimates are obtained across techniques with CTB data.

One point of interest in all of the estimates from Table 2.2, is the extent of dynamic consistency. Confirming recent findings with monetary payments when transaction costs and payment risk are closely controlled, we find virtually no evidence of present bias (Andreoni and Sprenger, 2012a; Gine, et al., 2012; Andersen et al., 2013; Augenblick, Niederle, and Sprenger, 2013; Kuhn, Kuhn, and Villeval, 2013). Across elicitation techniques and estimation strategies, the present bias parameter, β, is estimated close to one.

Substantial differences in estimates, particularly for utility function curvature, are obtained across the DMPL and the CTB. It is beyond the scope of this paper to provide a theoretical foundation for which elicitation is more likely to yield correct estimates. We instead take the approach that predictive validity is the relevant metric for assessing the value of each method. Our prediction exercises are considered next.

2.3.2 Predictive Validity

We consider predictive validity in two steps. First, we test within and between methods. That is, we examine the in- and out-of-sample fit for CTB and DMPL estimates on the CTB data. Correspondingly we examine the in- and out-of-sample fit for CTB and DMPL estimates on the DMPL data. Though one would expect the in-sample estimates to outperform the out-of-sample estimates, this exercise does yield one critical finding: the CTB estimates perform about as well out-of-sample as the DMPL estimates perform in-sample for intertemporal choices.

Second, we test strictly out-of-sample for both methods. We examine behavior in a BDM mechanism eliciting willingness to accept to relinquish a claim for $25 at a later date and two hypothetical measures for patience. These three out-of-sample environments are constructed such that model estimates generate point predictions for behavior. Hence, one can analyze differences between predicted and actual behavior and the correlation between the two. Importantly, in both exercises we account for individual heterogeneity by estimating discounting parameters for each individual separately (see appendix section 2.5.3 for details). For the CTB, individual level estimates are constructed based upon the estimation strategy of Table 2.2, Column (4). Individual level estimates of α, β and r are obtained for all 58 subjects.[60] For the DMPL, individual level estimates are constructed based upon the estimation strategy of Table 2.2, Column (3). Individual level estimates of α, β and r are obtained for all 58 subjects. These analyses demonstrate that CTB-based estimates outperform DMPL-based estimates in all three out-of-sample environments.[61]

2.3.2.1 Within and Between Methods

We begin by analyzing the CTB data. First, consider the in-sample fit for the CTB estimates. We use the parameter estimates from Table 2.2, column (4) to construct utilities for each option within a budget and compare the predicted utility-maximizing option to the chosen option. Using the aggregate CTB estimates, the predicted utility maximizing choice was chosen 45% of the time. Using individual CTB estimates, the predicted utility maximizing choice was chosen 75% of the time. Next, consider the out-of-sample fit for the DMPL estimates. We use the parameter estimates from Table 2.2, column (3) to construct utilities for each option within a budget and compare the predicted utility-maximizing option to the chosen option. Aggregate DMPL estimates predict 3% of CTB choices correctly and individual DMPL estimates predict 16% of CTB choices correctly.

The key out-of-sample failure for the DMPL estimates on the CTB data is gener-

[60]We opt to use the OLS estimates from Table 2.2, column (4), because individual level estimates are obtained for all 58 subjects. Using the NLS estimates of Table 2.2, column (5) very similar results are obtained, though the individual-level estimator converges for only 56 of 58 subjects.

[61]To account for estimation error, we also used the standard errors of the estimation to bootstrap the CTB and DMPL estimates for each person-choice combination. Since the results are quantitatively and qualitatively similar to those using the estimates alone, we do not report them here. One important dissimilarity, however, should be noted. When making DMPL predictions the bootstrapping procedure generates negative estimates of α in about 40% of the cases. If we exclude these, the predictive success of the bootstrapped individual level DMPL estimates is modestly better than the estimates alone. However, if we count these as incorrect predictions, the predictive success of the individual level DMPL estimates is reduced dramatically. Excluding negative α's skews the remaining α's toward 1, which we demonstrate below favors more accurate predictions.

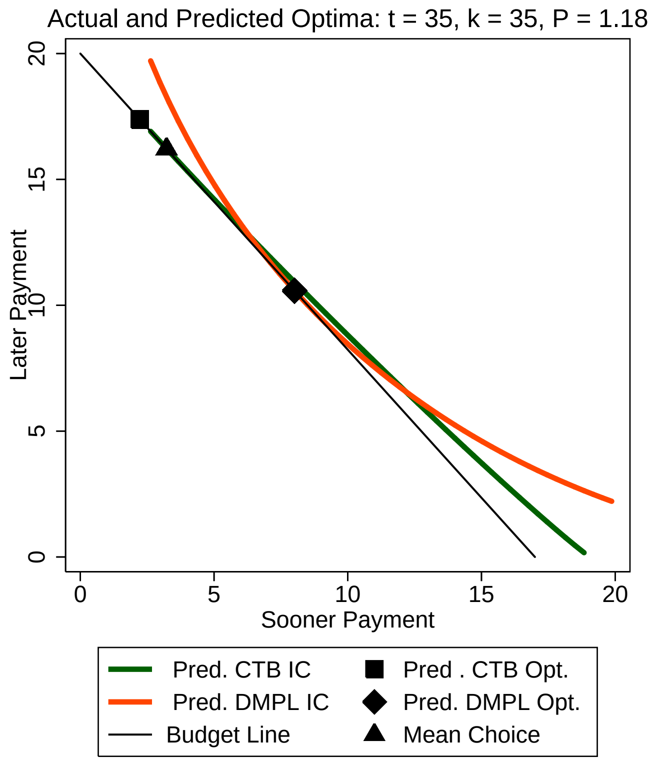


Figure 2.3: CTB and DMPL Prediction of CTB Data

ated by the high degree of estimated utility function curvature. Indeed, the majority of CTB choices are close to budget corners.[62] Figure 2.3 presents an example budget with corresponding predicted indifference curves and choices based on CTB and DMPL estimates. The high degree of curvature prevents the DMPL estimates from making corner predictions and hence leaves the estimates unable to match many data points.[63] Stated differently, the data themselves are a non-parametric rejection of the utility curvature estimate generated by the HL data, and the gains from the alternative source of curvature identification in the CTB swamps the bias incurred by using optimality to approximate corner solutions.

We perform an identical exercise for the DMPL data. We focus specifically on the

[62] To be specific 88 percent of CTB allocations are at one of the two budget corners. Additionally, 35 of 58 subjects have zero interior allocations, consistent with linear utility.

[63] See appendix section 2.5.5 for the the exercise conducted on all experimental budgets.

intertemporal MPL choices.[64] In-sample aggregate DMPL estimates predict 81% of MPL choices correctly and individual estimates predict 89% of MPL choices correctly. Interestingly, the CTB estimates perform almost as well out-of-sample as the DMPL estimates perform in-sample. Aggregate CTB estimates also predict 81% of MPL choices correctly and individual estimates predict 86% of MPL choices correctly.

From this exercise we note that using individual level estimates both estimation techniques perform well in-sample. However, the CTB estimates predict out-of-sample with greater accuracy than the DMPL estimates. In order to put the two methods on equal footing, we next consider the predictive ability of the techniques in environments where both sets of estimates are out-of-sample.

2.3.2.2 Pure Out-of-Sample

Following the experimental implementation of the CTB and DMPL, subjects were notified of their two payment dates, based on a randomly chosen experimental decision. We then elicited the amount they would be willing to accept in their sooner check instead of \$25 in the later check using a BDM technique with a uniform distribution of random prices drawn from $[\$15.00, \$24.99]$.[65] All 58 subjects from our estimation exercise provided a

[64]The HL data are considered in appendix section 2.5.6 and demonstrate, unsurprisingly that the DMPL estimates vastly outperform the CTB estimates on the HL data.

[65]Hence, stating a willingness to accept greater than or equal to \$25 implied a preference for the later payment in all states. Four subjects provided BDM bids of exactly \$25 and no subjects provided a BDM bid greater than \$25. Stating a willingness to accept lower than \$15 implied a preference for any sooner payment. No subjects provided a BDM bid less than \$15.

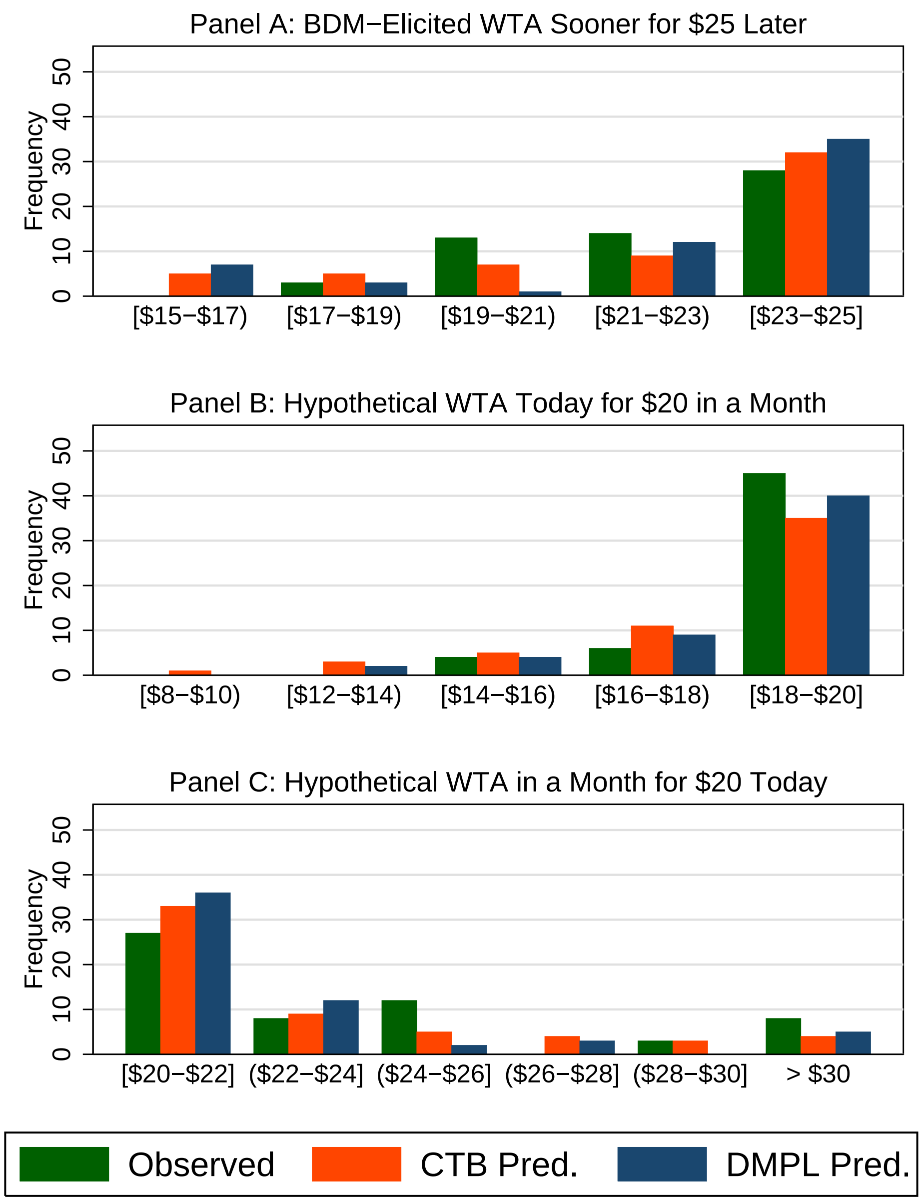


Figure 2.4: Out-of-Sample Distributions

BDM bid. The mean willingness to accept was $22.36 (s.d. $2.18). Figure 2.4, Panel A presents the distribution of willingness to accept BDM responses.

Based on the payment dates, we use the individual parameter estimates from the CTB and DMPL to predict subject responses. These predictions account for the fact that relevant payment dates may involve different values of t and k. Responses that are predicted to fall outside of the price bounds described above are top and bottom-coded, accordingly. The mean CTB based prediction is \$22.47 (s.d. \$3.09), while the mean DMPL prediction is \$22.48 (\$2.95). Tests of equality demonstrate that we fail to reject the null hypothesisof equal means between the true data and both our CTB and DMPL estimates, $(t_{57} = -0.247,\ p = 0.86)$, $(t_{57} = -0.251,\ p = 0.80)$, respectively. The predicted distributions from the CTB and DMPL estimates are also presented in Figure 2.4, Panel A. Though similar patterns to the true data emerge, Panel A does demonstrate some distributional differences, particularly at extreme values. Indeed, Kolmogorov-Smirnov tests of distributional equality reject the null hypothesis of equal distributions between observed and both CTB and the DMPL predictions, $(D = 0.414,\ p < 0.01)$, $(D = 0.241,\ p = 0.06)$, respectively. This suggests somewhat limited predictive validity at the distributional level.

Table 2.3, Panel A, columns (1) through (3) present tobit regressions analyzing the correlation between predicted and actual BDM behavior. In column (1) we show the CTB prediction to be significantly positively correlated with BDM bids. In contrast, an insignificant correlation is obtained in column (2) where the independent variable is the DMPL predicted bid. Further, in column (3) when both predictions are used in estimation, we find that DMPL predictions carry little explanatory power beyond that of the CTB. This indicates predictive validity of the CTB estimates, though not the DMPL estimates, at the individual level.

Table 2.3: Out-of-Sample Prediction

	CTB Predictions	DMPL Predictions	Both Predictions
	(1)	(2)	(3)
	Panel A: BDM-Elicited WTA Sooner for $25 Later		
CTB Prediction	0.230**		0.292**
	(0.094)		(0.118)
DMPL Prediction		0.079	-0.107
		(0.103)	(0.125)
Constant	17.273	20.658	18.310
	(2.124)	(2.339)	(2.433)
Pseudo R^2	0.023	0.002	0.026
N	58	58	58
	Panel B: Hypothetical WTA Today for $20 in One Month, $\$X_{today}$		
CTB Prediction	0.545***		0.465***
	(0.092)		(0.121)
DMPL Prediction		0.600***	0.158
		(0.129)	(0.164)
Constant	9.268	8.217	7.805
	(1.672)	(2.633)	(2.267)
Pseudo R^2	0.152	0.084	0.157
N	55	55	55
	Panel C: Hypothetical WTA in One Month for $20 Today, $\$Y_{month}$		
CTB Prediction	0.541*		0.956**
	(0.322)		(0.448)
DMPL Prediction		0.102	-0.987
		(0.535)	(0.736)
Constant	9.931	19.798	22.264
	(7.409)	(11.829)	(11.596)
Pseudo R^2	0.010	0.000	0.016
N	55	55	55

$^{***} \Rightarrow p < 0.01$, $^{**} \Rightarrow p < 0.05$, $^{*} \Rightarrow p < 0.10$. All correlation estimates are from tobit regressions of actual choices on individual-specific choice estimates generated from utility function parameters. The predicted choices are top and bottom-coded in the following way: Panel A top and bottom-coded at BDM price distribution bounds. Panel B top-coded at $20. Panel C bottom-coded at $20. Of the 58 subjects for whom we have parameter estimates and BDM bids, 3 are dropped from the hypothetical choice analysis. 2 of these 3 failed to provide survey responses for either hypothetical question and another is excluded due to extreme outlying DMPL predictions.

Our final two prediction exercises involve hypothetical data collected during the post-experiment survey. First, we asked subjects what amount of money, \$X$_{today}$, today would make them indifferent to \$20 in a month. Second, we asked subjects what amount of money, \$Y$_{month}$, in a month would make them indifferent to \$20 today. Both measures are noisy with subjects at times answering free-form.[66] 56 of 58 subjects from our estimation exercise provided values for \$X$_{today}$ and \$Y$_{month}$. Figure 2.4, Panels B and C present these data. The data for \$X$_{today}$ are top-coded at \$20 while the data for \$Y$_{month}$ are bottom-coded at \$20. Following an identical strategy to that above, Panels B and C also present the distribution of responses predicted from CTB and DMPL individual estimates, top and bottom-coded accordingly. One subject's DMPL estimates produced a predicted value of \$Y$_{month}$ in excess of \$1,000 and a \$X$_{today}$ value of approximately \$0. Excluding this outlier, our analysis focuses on 55 subjects. In nearly all cases, we reject the null hypothesis of equal means between predicted values and actual values.[67] Further, distributional tests frequently reject the null hypothesis of equality suggesting limited predictive validity at the distributional level.[68]

[66]In the first question, one subject responded 'Any amount over \$20'. This response was coded as \$20. This subject gave the same response in the second question and was again coded as \$20. In the second question, one subject responded, '\$19.05 plus one dollar in a month'. This was coded as \$20.05.

[67]The mean actual value of \$X$_{today}$ is \$18.79 (s.d. \$1.50). The CTB-based prediction for \$X$_{today}$ is \$18.29 (s.d. \$2.36). The DMPL-based prediction for \$X$_{today}$ is \$18.44 (s.d. \$1.76). We reject the null hypothesis of equal means between the true data and our CTB estimates ,though not our DMPL estimates, $(t_{54} = 2.13,\ p = 0.04)$, $(t_{54} = 1.63,\ p = 0.11)$, respectively. The mean actual value of \$Y$_{month}$ is \$24.27 (s.d. \$6.62). The CTB-based prediction for \$Y$_{month}$ is \$22.35 (s.d. \$3.86). The DMPL-based prediction for \$Y$_{month}$ is \$21.92 (s.d. \$2.46). We reject the null hypothesis of equal means between the true data and both our CTB and DMPL estimates, $(t_{54} = 2.04,\ p = 0.05)$, $(t_{54} = 2.48,\ p = 0.02)$, respectively.

[68]The KS statistic for the comparison of \$X$_{today}$ across the true data and the CTB prediction is $D = 0.184,\ (p = 0.25)$. For the comparison of \$X$_{today}$ across the true data and the DMPL prediction is is $D = 0.222,\ (p = 0.10)$. The KS statistic for the comparison of \$Y$_{month}$ across the true data and the CTB prediction is $D = 0.207,\ (p = 0.14)$. For the comparison of \$Y$_{month}$ across the true data and the DMPL prediction is is $D = 0.259,\ (p = 0.04)$.

When considering the extent of correlations at the individual level, a different conclusion is drawn. Table 2.3, Panels B and C present tobit regressions similar to Panel A, where the dependent variable is either \$$X_{today}$ or \$$Y_{month}$. Again we find the CTB predictions to carry significant correlations with the true measures. Though in Panel B, the DMPL prediction does significantly correlate with observed behavior, the DMPL predictions provide limited added predictive power beyond the CTB predictions. This again indicates predictive validity of the CTB estimates at the individual level.

Across our three out-of-sample exercises we find that both the CTB and DMPL can mis-predict, at times importantly, the distribution of behavior. However, at the individual level predictive validity is apparent, particularly for CTB-based estimates. DMPL-based estimates at times provide little independent and additional predictive power in our out-of-sample environments.

2.3.2.3 Probabilistic Choice and Multiple Switching

While all of the predictions discussed above were generated via utility maximization, conditional on parameter values, the Luce model strategy suggests that another way of doing so would be to use a utility index with a decision error parameter to construct choice probabilities. This decision error allows one to connect preferences to choice probabilities via some functional form.[69] The aggregate in-sample fit of these models (estimated via maximum likelihood) may be very good but the out-of-sample prediction may falter. This may be for reasons of the parameter estimates being inapplicable or due to the assumption

[69]The specific functional form comes from the assumed random utility model and error distribution.

of probabilistic choice itself. In the case of price lists, a lot of decision error means a lot of multiple switching because the decision error is agnostic with regard to preferences.

Therefore, another way of asking whether the estimates faithfully describe the data is to consider the degree of randomness in choice exhibited and the degree predicted. Of the 64 subjects who took part in the experiment, none exhibited multiple-switching behavior in the MPL task.[70] However, the Luce probabilistic choice model used to estimate the DMPL parameters (and probabilistic choice models generally) predicts choice probabilities that necessarily allow for switching more than once with some non-zero probability. We simulate 1000 sets of our MPL data using these predicted choice probabilities and find that the DMPL parameters and Luce model predicts that 86% of subjects should exhibit an “irrational” switch.

How much of this gap is due to the model and how much is due to the parameters themselves? To determine this, we run the CTB parameters through the Luce model, borrowing the DMPL estimate of ν, to again simulate 1000 sets of our data. This exercise predicts that 57% of subjects should exhibit an “irrational” switch. Given that a curvature parameter away from 1 directly attenuates utility differences between options, the CTB-DMPL gap make sense. The remaining Data-CTB gap is due to the Luce model itself; there are no hallmarks of probabilistic choice in the data.

[70] All exclusions for multiple-switching were from violations on the HL task.

2.4 Conclusion

We compare two recent innovations for the experimental identification and estimation of time preferences, the Convex Time Budget (CTB) of Andreoni and Sprenger (2012a) and the Double Multiple Price List (DMPL) of Andersen et al. (2008). Both innovations focus on generating measures of discounting which are not confounded by utility function curvature. The primary avenue along which the methods are compared is predictive validity. We examine the extent to which estimated utility parameters can predict behavior across experimental methods and in out-of-sample environments. At the distributional level, we find that both methods make predictions close to average behavior, though they often miss key elements of the distribution. At the individual level, we find CTB-based estimates to have increased predictive power relative to DMPL estimates.

Our experiment suggests one prominent explanation for the observed differences between CTB and DMPL-based estimates. The key distinction in identification strategies across the two methods comes from the source of information for utility function curvature. In the CTB, curvature is informed by sensitivity to changing interest rates. In the DMPL, curvature is informed by risky choice. The most striking difference in parameter estimates across the two methods is the level of curvature. It is beyond the scope of this paper to state definitively which source of identification is correct. However, if one is interested in predicting intertemporal choice, our results demonstrate that the latter generates less informative estimates. Importantly, this is precisely because CTB data consist mostly of corner solutions. We wish to emphasize that this aspect of the CTB approach is in fact the critical feature of the data that is responsible for such substantial improvements in predictive validity over the DMPL, and in particular, over using the HL device to identify curvature.

In motivating our study we suggested predictive power as a primary metric of success. We take the first step in this direction by exploring out-of-sample choices of our subjects made in the experiment. An essential test that remains, however, is to use these measurements of time preference to predict behavior outside of the experiment. In addition to laboratory refinement of the techniques presented here, a key next step is expanding to target populations for whom extra-lab choices are observable. Linking precisely measured discounting parameters to important intertemporal decisions is a promising avenue of future research.

2.5 Acknowledgements

Chapter 2, in full, has been submitted for publication of the material at the *American Economic Journal: Microeconomics*. Andreoni, James; Kuhn, Michael A.; Sprenger, Charles. The dissertation author was the primary investigator and author of this material.

2.6 Appendix

In this Appendix, we provide a full description of the ICT identification, Luce stochastic error model identification, individual level parameter estimates and their cross-method correlations, estimates with different background consumption level specifications, data on the CTB predictions, an analysis of HL prediction, and the experimental forms.

2.6.1 ICT Identification

Our data is notably different from AS in that we offer only six discrete options along a budget, whereas they offer 101. This means that the Euler and demand equations do not hold exactly at the points elicited from our experiment. If the differences between optima and choices depend systematically on the independent variables, this could bias our results. One way to think of this problem is as non-classical measurement error on the dependent variable.[71] As a check against this potential problem, we ignore the cardinal information associated with our observed responses and treat them as ordinal indicators of preference. We assume that optimality holds only for the underlying, unobserved optimal choices from fully-convex budgets, and that our observed data are related only probabilistically to the optimality conditions, but not subject to the same identification condition. The key feature that distinguishes this approach from techniques like the Luce stochastic error model or multinomial logit is that we maintain the assumption of optimality and thus the ordering of the choice options.

Our starting point for the ICT is a simplified version of (2), the OLS regression

[71] Which can also be expressed as an omitted variable bias.

equation. Indexing the variables by i for individual and j for budget number, we have

$$z^*_{ij} = ln\left(\frac{x^*_{t(ij)}}{x^*_{t+k(ij)}}\right) = \gamma_1 t_{0ij} + \gamma_2 k_{ij} + \gamma_3 ln(P_{ij}) + e_{ij}, \tag{5}$$

where the starred variables indicate the underlying, unobserved optima. We can order all 6 choices along a budget in terms of their preference for sooner payment: call these $c = 1, 2, ...6$. We define the following correspondence between z^* and c:

$$c = \begin{cases} 1 & \text{if } z^* > K^1 \\ 2 & \text{if } K^1 > z^* > K^2 \\ \vdots & \vdots \\ 6 & \text{if } K^5 > z^* \end{cases}$$

The cut points, $K_1...K_5$, should not be interpreted as points of indifference between the adjacent choices, because conditional on parameter values, there is no indifference between adjacent choices. They are features of both the observed and unobserved parts of preferences. If they are known, it is straightforward to construct choice probabilities by making a distributional assumption on the error term. For $e_{ij} \sim N(0, \sigma^2)$, we have,

$$Pr(c_{ij} = n) = Pr(K^{n-1}_j < z^*_{ij} < K^n_j) =$$

$$Pr(K^{n-1}_j - \gamma_1 t_{0ij} - \gamma_2 k_{ij} - \gamma_3 ln(P_{ij}) < e_{ij} < K^n_j - \gamma_1 t_{0ij} - \gamma_2 k_{ij} - \gamma_3 ln(P_{ij}))$$

$$= \Phi\left(\frac{K_j^{n-1}}{\sigma} - \frac{\gamma_1}{\sigma}t_{0ij} - \frac{\gamma_2}{\sigma}k_{ij} - \frac{\gamma_3}{\sigma}ln(P_{ij})\right) - \Phi\left(\frac{K_j^{n}}{\sigma} - \frac{\gamma_1}{\sigma}t_{0ij} - \frac{\gamma_2}{\sigma}k_{ij} - \frac{\gamma_3}{\sigma}ln(P_{ij})\right), \tag{6}$$

where Φ represents the standard normal CDF. This holds exactly for the all interior choice options and the derivation for the corner solution probabilities follows the same logic. We estimate the cut points simultaneously with the other parameters using maximum likelihood.

Note that (6) demonstrates the γ parameters are only identified up to a constant of proportionality (σ) in this model, as are the cut points. Unfortunately, this prevents us from precisely estimating α because $\gamma_3 = \frac{1}{\alpha-1}$. The estimate of $\alpha = \frac{\sigma}{\gamma_3} + 1$ is thus directly affected by this lack of identification. However $\gamma_1 = \frac{ln(\beta)}{\alpha-1}$, implying $\beta = exp(\frac{\gamma_1}{\gamma_3})$ and $\gamma_2 = \frac{ln(\delta)}{\alpha-1}$, implying $\delta = exp(\frac{\gamma_2}{\gamma_3})$. Because these two utility parameters are identified from ratios of the γ coefficients, the constant of proportionality does not affect the estimates. Examining whether these parameter estimates differ across methods serves as a robustness check on the OLS and NLS procedures against the potential non-standard measurement error bias introduced by ignoring the interval nature of the data in those approaches.

Note that in the expression above the cutoffs are indexed by decision, j. Ideally, we would want to identify all five cutoffs specific to all 24 budgets, but to maintain statistical feasibility we make an assumption that reduces the cut point estimation problem from 120 to 5 parameters. However, it is important that the assumption we make allows the cut points to vary across budgets to reflect price and income changes. Note that the error, e_{ij} is in units of the log consumption ratio. Using this fact, we assume that the cut point between choices n and $n-1$ is defined as the log of a linear combination of the consumption ratios at choices n and $n-1$ according to mixing parameter $\lambda^n \in [0,1]$. To state this formally, define K_j^n as the cut point that determines whether and individual selects option n or $n-1$ on choice j. Then

$$K_j^n = ln\left(\frac{x_{t(j)}(c_j = n)}{x_{t+k(j)}(c_j = n)}\lambda^n + \frac{x_{t(j)}(c_j = n-1)}{x_{t+k(j)}(c_j = n-1)}(1-\lambda^n)\right). \tag{7}$$

Assumption: $\lambda_j^n = \lambda_{j'}^n \quad \forall \quad (n, j, j') \in (\{j = 1...24\}, \{j' = 1...24\}, \{n = 1...5\})$.

While the mixing parameters for each interval are constant across budgets, the actual cut points associated with them adjust for the different properties of each budget.

There are other similar approaches to the ICT that one could take in our case. For example, an essentially identical exercise could be performed using the demand function rather than the tangency condition. However, the non-linearity of this function combined with the necessity of estimating cut-points makes the likelihood function very poorly behaved. More standard approaches would involve random utility models that do not take advantage of optimality conditions.

2.6.2 Luce Stochastic Error Model Identification

AHLR use choice probabilities based on HL's adaptation of work by Luce (1959) to construct a likelihood function. Recall that according to this model, if an individual is presented with options X and Y, their probability of choosing option X is

$$Pr(c = X) = \frac{U(X)^{\frac{1}{\nu}}}{U(X)^{\frac{1}{\nu}} + U(Y)^{\frac{1}{\nu}}}.$$

ν represents deviations from deterministic choice. In the context of intertemporal choice, assume X represents sooner income and Y represents later income. Risk decisions from the HL are modeled similarly. For options L and R, the probability of choosing L is

$$Pr(c = L) = \frac{U(L)^{\frac{1}{\mu}}}{U(L)^{\frac{1}{\mu}} + U(R)^{\frac{1}{\mu}}}.$$

Every individual decision on both the risk and time tasks generates one entry in the log-likelihood function. We use s to denote the risk decision index, j to denote the time decision index and i to denote individuals. The risk and time decisions enter the global DMPL likelihood function under an independence assumption that maintains complete linearity. This yields a log-likelihood function of

$$L = \sum_{ij} 1(c_{ij} = X_j) ln\left(\frac{U(X_j)^{\frac{1}{\nu}}}{U(X_j)^{\frac{1}{\nu}} + U(Y_j)^{\frac{1}{\nu}}}\right) + \sum_{ij} 1(c_{ij} = Y_j) ln\left(\frac{U(Y_j)^{\frac{1}{\nu}}}{U(X_j)^{\frac{1}{\nu}} + U(Y_j)^{\frac{1}{\nu}}}\right) +$$

$$\sum_{is} 1(c_{is} = R_s) ln\left(\frac{U(R_s)^{\frac{1}{\mu}}}{U(R_s)^{\frac{1}{\nu}} + U(L_s)^{\frac{1}{\nu}}}\right) + \sum_{is} 1(c_{is} = L_s) ln\left(\frac{U(L_s)^{\frac{1}{\mu}}}{U(R_s)^{\frac{1}{\nu}} + U(L_s)^{\frac{1}{\nu}}}\right). \quad (8)$$

2.6.3 Individual Parameter Estimates

Table A2.1: Individual-Specific Parameter Estimates

	Median	Mean	Standard Deviation	10th Pctile.	90th Pctile.
CTB					
α	0.937	0.936	0.030	0.915	0.966
β	1.084	1.060	0.160	0.839	1.174
r	0.692	33.553	197.117	-0.880	7.454
DMPL					
α	0.488	-0.178	3.426	0.231	0.958
β	0.995	2.320	9.947	0.948	1.027
r	0.282	0.994	2.649	-0.023	2.493

Estimates are obtained using OLS for the CTB and the Luce stochastic error model for the DMPL. There are 58 observations of each parameter.

Table A2.1 presents the individual-specific utility parameter estimates. The medians correspond closely to the aggregate estimates presented in Section 2.3.1. Using these measures, we can look at the between-method correlation for each parameter. Importantly, there are no significant pairwise correlation between measures of curvature, present-bias and discount rate across the two methods, ($\rho = 0.046,\ p = 0.773$), ($\rho = -0.073,\ p = 0.588$), ($\rho = 0.067,\ p = 0.619$), respectively.

2.6.4 Background Parameter Specifications

For the main analysis we consider the experimental allocations in a vacuum. However, the degree to which laboratory sensitivity to stakes depends on extra-laboratory income and consumption is unresolved. Furthermore, all subjects were provided with a $10 show-up fee that was divided into two payments of $5 and split between the two payment

Table A2.2: Aggregate Utility Parameter Estimates with Show-up Fees

	Discounting	Curvature	Discounting and Curvature			
Elicitation:	MPL	HL	DMPL		CTB	
Estimation:	ML	ML	ML	OLS	NLS	ICT
	(1)	(2)	(3)	(4)	(5)	(6)
Utility Parameters						
r	0.737		0.456	0.658	0.828	0.795
	(0.148)		(0.096)	(0.371)	(0.228)	(0.245)
β	0.989		0.992	1.017	0.998	0.999
	(0.008)		(0.006)	(0.021)	(0.014)	(0.018)
α		0.080	0.083	0.674	0.784	$0.831^{\dagger}$
		(0.092)	(0.091)	(0.018)	(0.011)	(0.023)
Error Parameters						
ν	0.065		0.003			
	(0.007)		(0.003)			
μ		0.009	0.009			
		(0.010)	(0.010)			
Clusters	58	58	58	58	58	58
N	1392	1160	2552	1392	1392	1392
Log Likelihood	-545	-326	-871			-1514
R^2				0.420	0.536	

†: The ICT estimate for α is only identified up to a constant. See Appendix A.1 for details.
Standard errors clustered at the individual level in parentheses. Each individual made 20 decisions on the HL, 24 decision on the MPL (and therefore 44 decisions on the DMPL) and 24 decisions on the CTB. In columns (1) through (3) HL, MPL and DMPL estimates are obtained via maximum likelihood using Luce's (1959) stochastic error probabilistic choice model. The CTB is estimated in three different ways: ordinary least squares (OLS) using the Euler equation (2), non-linear least squares (NLS) using the demand function (3) and interval-censored tobit (ICT) maximum likelihood using the Euler equation (2). All maximum likelihood models are estimated using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm.

dates.

Shifting income levels in both periods affects the levels of our estimates, but here we demonstrate that alternative specifications do not yield different qualitative results. Table A2.2 replicates our main Table of results in Section 2.3.1 with the $5 payments added

Table A2.3: Actual and Predicted Optima on CTB Data

t	k	P	CTB Opt.	DMPL Opt.	Mean Choice	Median Choice
0	35	1.05	9.00	9.68	8.91	7.60
		1.11	4.60	8.87	7.08	0.00
		1.18	1.88	8.08	4.92	0.00
		1.25	0.65	7.31	2.87	0.00
		1.43	0.05	5.84	1.01	0.00
		1.82	0.00	3.83	0.23	0.00
0	63	1.00	16.66	10.83	19.52	20.00
		1.05	12.67	9.99	10.16	13.30
		1.18	3.68	8.36	5.63	0.00
		1.33	0.43	6.81	2.59	0.00
		1.67	0.01	4.65	0.95	0.00
		2.22	0.00	2.78	0.09	0.00
35	35	1.05	9.91	9.59	7.60	0.00
		1.11	5.29	8.79	5.71	0.00
		1.18	2.22	8.01	3.22	0.00
		1.25	0.78	7.24	2.10	0.00
		1.43	0.07	5.78	0.92	0.00
		1.82	0.00	3.78	0.53	0.00
35	63	1.00	17.17	10.74	19.17	20.00
		1.05	13.46	9.91	10.29	17.10
		1.18	4.26	8.29	4.98	0.00
		1.33	0.52	6.74	2.54	0.00
		1.67	0.01	4.60	0.87	0.00
		2.22	0.00	2.74	0.37	0.00

In each of the 24 rows, we use observed data from the 58 individuals who comprised our estimation sample. Optima are calculated using aggregate estimates and are all in terms of sooner income.

to each time period. We document substantial sensitivity in discounting and curvature estimates, particularly for the DMPL. Importantly, the difference in curvature across methods remains pronounced.

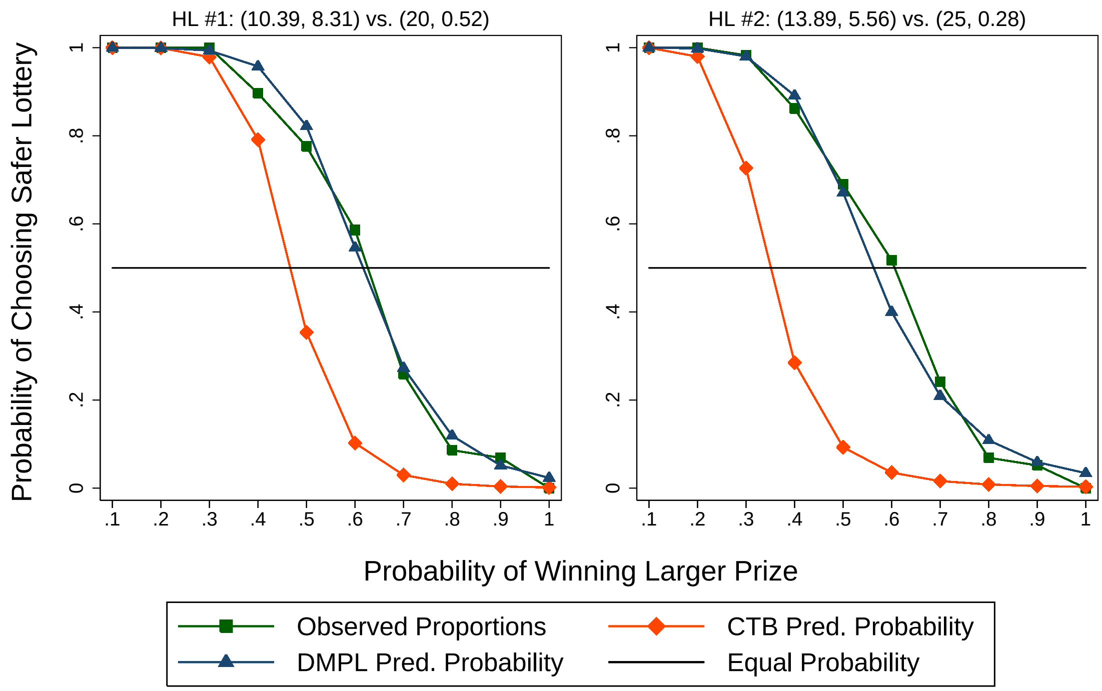


Figure A2.1: HL Prediction Exercise

2.6.5 CTB Prediction Data

In Table A2.3, we present predicted optima and observed CTB choice based on CTB and DMPL estimates.

2.6.6 HL Predictions

A method of testing whether or the risky and riskless data generate conformable measures of concavity is to use the α parameters as estimated from the CTB to try and predict risky choices. Both aggregate and individual CTB estimates predict 82% of HL choices correctly. By comparison both aggregate and individual DMPL estimates predict with 90% accuracy.[72]

[72] The difference is statistically significant with $p = 0.005$. Standard errors are clustered by individual.

Figure A2.1 plots the HL choice probabilities[73] for each measure of curvature and observed choices for each of our HL tasks. This illustrates that the CTB fails to predict enough risk-aversion to explain the data.[74]

2.6.7 Experimental Stimuli

We provide the following stimuli (in this order): explanation of payment method, general instructions, CTB instructions, MPL instructions, HL instructions, BDM instructions.

[73]These are calculated using the Luce Stochastic Error model. In the case of the CTB estimates, we borrow the value fo μ from the DMPL estimation.

[74]Testing for equality of the predicted probabilities rejects with $p < 0.001$, standard errors clustered by individual.

Welcome

Welcome and thank you for participating

Eligibility for this study

To be in this study, you need to meet these criteria.

You must have a campus mailing address of the form:

```
YOUR NAME
9450 GILMAN DR 92(MAILBOX NUMBER)
LA JOLLA CA 92092-(MAILBOX NUMBER)
```

Your mailbox must be a valid way for you to receive mail from now through the end of the Spring Quarter.

You must be willing to provide your name, campus mail box, email address, and student PID. This information will only be seen by Professor Andreoni and his assistants. After payment has been sent, this information will be destroyed. Your identity will not be a part of any subsequent data analysis.

You must be willing to receive your payment for this study by check, written to you by Professor James Andreoni, Director of the UCSD Economics Laboratory. The checks will be drawn on the USE Credit Union on campus. This means that, if you wish, you can cash your checks for free at the USE Credit Union any weekday from 9:00 am to 5:00 pm with valid identification (driver's license, passport, etc.).

The checks will be delivered to you at your campus mailbox at a date to be determined by your decisions in this study, and by chance. The latest you could receive payment is the last week of classes in the Spring Quarter.

If you do not meet <u>all of these criteria</u>, please inform us of this now.

Instructions

Earning Money

To begin, you will be given a $10 thank-you payment, just for participating in this study! You will receive this thank-you payment in two equally sized payments of $5 each. The two $5 payments will come to you at two different times. These times will be determined in the way described below.

In this study, you will make 49 choices over how to allocate money between two points in time, one time is "earlier" and one is "later." Both the earlier and later times will vary across decisions. This means you could be receiving payments as early as today, and as late as the last week of classes in the Spring Quarter, or possibly two other dates in between.

Once all 49 decisions have been made, we will **randomly select one of the 49 decisions as the decision-that-counts**. We will use the decision-that-counts to determine your actual earnings. Note, since all decisions are equally likely to be chosen, you should make each decision as if it will be the decision-that-counts.

When calculating your earnings from the decision-that-counts, we will add to your earnings the two $5 thank you payments. Thus, you will always get paid at least $5 at the chosen earlier time, and at least $5 at the chosen later time.

IMPORTANT: All payments you receive will arrive to your campus mailbox. That includes payments that you receive today as well as payments you may receive at later dates. On the scheduled day of payment, a check will be placed for delivery in campus mail services by Professor Andreoni and his assistants. *Campus mail services guarantees delivery of 100% of your payments by the following day.*

As a reminder to you, the day before you are scheduled to receive one of your payments, we will send you an e-mail notifying you that the payment is coming.

On your table is a business card for Professor Andreoni with his contact information. Please keep this in a safe place. If one of your payments is not received you should immediately contact Professor Andreoni, and we will hand-deliver payment to you.

Your Identity

In order to receive payment, we will need to collect the following pieces of information from you: name, campus mail box, email address, and student PID. This information will only be seen by Professor Andreoni and his assistants. After all payments have been sent, this information will be destroyed. Your identity will not be a part of subsequent data analysis.

You have been assigned a participant number. This will be linked to your personal information in order to complete payment. After all payments have been made, only the participant number will remain in the data set.

On your desk are two envelopes: one for the sooner payment and one for the later payment. Please take the time now to address them to yourself at your campus mail box.

NAME: ______________________________ PID:____________________

How It Works:

In the following four sheets you are asked to make 24 decisions involving payments over time. Each row on the sheets is a decision and is numbered from 1 to 24.

Each row will feature a series of options. Each option consists of a sooner payment AND a later payment. You are asked to pick your favorite option in each row by checking the box below it. You should pick the combination of sooner payment AND later payment that you like the most. For each row, mark only one box.

Here is an example row:

1.	payment TODAY …	$19.00	$15.20	$11.40	$7.60	$3.80	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐

In this example, you are asked to choose your favorite combination of payment today AND payment in 5 weeks. As you can see, the sooner payment varies in value from $19 to $0 and the later payment varies in value from $0 to $20.

Note that there is a trade-off between the sooner payment and the later payment across the options. As the sooner payment goes down, the later payment goes up.

If someone's favorite option were $19.00 today AND $0 in five weeks, they would mark as follows:

1.	payment TODAY …	$19.00	$15.20	$11.40	$7.60	$3.80	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐

If someone's favorite option were $0 today AND $20 in five weeks, they would mark as follows:

1.	payment TODAY …	$19.00	$15.20	$11.40	$7.60	$3.80	$0
	and payment in 5 WEEKS	$0	$4.00	$8.00	$12.00	$16.00	$20.00
		☐	☐	☐	☐	☐	☐

How to proceed:
There are 4 sheets, each with 6 decisions, making 24 decisions in total. Each decision has a number from 1 to 24.

NUMBERS 1 THROUGH 6: Each option has one payment today AND one payment in 5 weeks.
NUMBERS 7 THROUGH 12: Each option has one payment today AND one payment in 9 weeks.
NUMBERS 13 THROUGH 18: Each option has one payment in 5 weeks AND one payment in 10 weeks.
NUMBERS 19 THROUGH 24: Each option has one payment in 5 weeks AND one payment in 14 weeks.

Your decisions represent 24 of the 49 choices you make in the experiment. If, after all 49 choices are made, a number from 1 to 24 is drawn, these sheets will determine your payoffs. This number will determine which decision (from 1 to 24) will determine your payoffs. The sooner and later payments from the option you choose in the decision-that-counts will be added to your sooner and later $5 thank-you payments.

Remember that each decision could be the decision-that-counts! It is in your interest to treat each decision as if it could be the one that determines your payment.

NAME: ______________________________ PID:____________________

How It Works:

In the following four sheets you are asked to make 24 decisions involving payments over time. Each row on the sheets is a decision and is numbered from 1 to 24.

Each row will feature a series of options. Each option consists of a sooner payment AND a later payment. You are asked to pick your favorite option in each row by checking the box below it. You should pick the combination of sooner payment AND later payment that you like the most. For each row, mark only one box.

Here is an example row:

1.	payment TODAY …	$19.00	$0
	and payment in 5 WEEKS	$0	$20.00
		☐	☐

In this example, you are asked to choose your favorite combination of payment today AND payment in 5 weeks. As you can see, the sooner payment varies in value from $19 to $0 and the later payment varies in value from $0 to $20.

Note that there is a trade-off between the sooner payment and the later payment across the options. As the sooner payment goes down, the later payment goes up.

If someone's favorite option were $19.00 today AND $0 in five weeks, they would mark as follows:

1.	payment TODAY …	$19.00	$0
	and payment in 5 WEEKS	$0	$20.00
		☐	☐

If someone's favorite option were $0 today AND $20 in five weeks, they would mark as follows:

1.	payment TODAY …	$19.00	$0
	and payment in 5 WEEKS	$0	$20.00
		☐	☐

How to proceed:
There are 4 sheets, each with 6 decisions, making 24 decisions in total. Each decision has a number from 1 to 24.

NUMBERS 1 THROUGH 6: Each option has one payment today AND one payment in 5 weeks.
NUMBERS 7 THROUGH 12: Each option has one payment today AND one payment in 9 weeks.
NUMBERS 13 THROUGH 18: Each option has one payment in 5 weeks AND one payment in 10 weeks.
NUMBERS 19 THROUGH 24: Each option has one payment in 5 weeks AND one payment in 14 weeks.

Your decisions represent 24 of the 49 choices you make in the experiment. If, after all 49 choices are made, a number from 1 to 24 is drawn, these sheets will determine your payoffs. This number will determine which decision (from 1 to 24) will determine your payoffs. The sooner and later payments from the option you choose in the decision-that-counts will be added to your sooner and later $5 thank-you payments.

Remember that each decision could be the decision-that-counts! It is in your interest to treat each decision as if it could be the one that determines your payment.

NAME: ______________________________________ PID:______________________

<u>How It Works:</u>

In the following two sheets you are asked to choose between options: Option A or Option B. On each sheet you will make ten choices, one on each row. For each decision row you will have to choose either Option A or Option B. You make your decision by checking the box next to the option you prefer more. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order.

There are a total of 20 decisions on the following sheets. The sheets represent one of the 47 choices you make in the experiment. If the number 46 is drawn, these sheets will determine your payoffs. If the number 46 is drawn, a second number will also be drawn from 1 to 20. This will determine which decision (from 1 to 20) on the sheets is the decision-that-counts. The option you choose (either Option A or Option B) in the decision-that-counts will then be played. You will receive your payment from the decision–that-counts immediately. Your $5 sooner and later thank-you payments, however, will still be mailed as before. The sooner payment will be mailed today and the later payment will be mailed in 5 weeks.

Playing the Decision-That-Counts:

Your payment in the decision-that-counts will be determined by throwing a 10 sided die. Now, please look at Decision 1 on the following sheet. Option A pays $10.39 if the throw of the ten sided die is 1, and it pays $8.31 if the throw is 2-10. Option B yields $20 if the throw of the die is 1, and it pays $0.52 if the throw is 2-10. The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between $10.39 or $20.

Remember that each decision could be the decision-that-counts! It is in your interest to treat each decision as if it could be the one that determines your payoff.

NAME: __________________________ PID:______________________

Thank you for participating. Please write the dates that you are scheduled to receive your sooner and later payments:

sooner date __________ AND later date __________

As an additional thank-you, one person will be chosen to receive an additional $25 in their later check. This will be decided randomly and each of you has an equal chance of receiving the $25. If you are chosen to receive the $25 later, you will also have the possibility of taking a smaller amount in your earlier check. All you have to do is tell us the lowest amount you would accept in your sooner check to compensate you for not getting the $25 in your later check.

<u>How it works:</u> In the box below, write the smallest amount that you'd be willing to accept in your sooner check *in exchange for* the additional $25 in your later check. That is, you should state the amount of money paid to you at the sooner date that would be just as good to you as being paid $25 at the later date.

<u>Why should you tell the truth?</u> Once you write the amount in the box below we will collect all of the sheets. Then we will randomly pick a number between $15.00 and $24.99. All numbers from $15.00 to $24.99 are equally likely to be chosen. If the randomly chosen number is *greater than* the amount you write *you will be paid the <u>random number</u>* at the *sooner date*. If the randomly chosen number is *smaller than* the amount you write *you will be paid <u>$25</u>* at the *later date*.

<u>Example:</u> Let's say someone writes $21.80 in the box below. If the random number is $19.40, they receive the $25 later. If the random number is $22.49, they receive $22.49 sooner.

<u>What if I don't write my true value?</u> Consider the following stories.

Story 1: Suppose that, in truth, Person A is really indifferent between $22.85 paid sooner and $25 paid later. Instead of writing $22.85, Person A writes a lower number, say $20.00. Then, the random number is drawn and it winds up being $21.50. Person A receives $21.50 sooner and not $25 later. Person A is disappointed. $21.50 sooner is worse than $22.85 sooner; and $22.85 sooner is worth the same as $25 later. So the $21.50 is worse than the $25 Person A could have had by writing his true value. *By writing his true value instead of a lower number, Person A would be better off.*

Story 2: Suppose that, in truth, Person B is really indifferent between $19.65 paid sooner and $25 paid later. Instead of writing $19.65, Person B writes a higher number, say $23.25. Then, the random number is drawn and it winds up being $21.50. Person B receives the $25 later. Person B is disappointed. $25 later is worth the same as $19.65 sooner. $19.65 sooner is less than the $21.50 Person B could have had by writing her true value. *By writing her true value instead of a higher number, Person B would be better off.*

As you can see, the best idea is to write down your true value. Not a penny more and not a penny less!

Please write the smallest amount you'd truly be willing to accept in your sooner check in exchange for the additional $25 in your later check.

I am indifferent between $________ in my sooner check and $25 in my later check.

2.7 References

Steffen Andersen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. Eliciting risk and time preferences. *Econometrica*, 76(3):583–618, 2008.

Steffen Andersen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. Discounting behavior: A reconsideration. Working Paper, Center for the Economic Analysis of Risk, 2013.

James Andreoni and Charles Sprenger. Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–3356, 2012a.

James Andreoni and Charles Sprenger. Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–3376, 2012b.

Dan Ariely, George Loewenstein, and Drazen Prelec. Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118(1):73–106, 2003.

Nava Ashraf, Dean Karlan, and Wesley Yin. Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics*, 121 (2):635–672, 2006.

Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over time: Dynamic inconsistency in real effort tasks. NBER Working Paper No. 18734, 2013.

Gordon M. Becker, Morris H. DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232, 1964.

Marco Cagetti. Wealth accumulation over the life cycle and precautionary savings. *Journal of Business and Economic Statistics*, 21(3):339–353, 2003.

Christopher F. Chabris, David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky. Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, 37(2):237–269, 2008a.

Christopher F. Chabris, David Laibson, and Jonathon P. Schuldt. Intertemporal choice. In S. Durlauf and L. Blume, editors, *New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008b.

Maribeth Coller and Melonie B. Williams. Eliciting individual discount rater. *Experimental Economics*, 2:107–127, 1999.

Robin P. Cubitt and Daniel Read. Can intertemporal choice experiments elicit preferences for consumption? *Experimental Economics*, 10(4):369–389, 2007.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260, 2010.

Shane Frederick, George Loewenstein, and Ted O'Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401, 2002.

Xavier Gine, Jessica Goldberg, Dan Silverman, and Dean Yang. Revising commitments: Field evidence on the adjustment of prior choices. NBER Working Paper No. 18065, 2012.

Pierre-Olivier Gourinchas and Jonathan A. Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002.

Glenn W. Harrison, Morten I. Lau, and Melonie B. Williams. Estimating individual discount rates in denmark: A field experiment. *American Economic Review*, 92(5):1606–1617, 2002.

Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. Identifying time preferences with experiments: Comment. Working Paper, Center for the Economic Analysis of Risk, 2013.

Jerry A. Hausman. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics*, 10(1):33–54, 1979.

Charles A. Holt and Susan K. Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

Michael A. Kuhn, Peter Kuhn, and Marie Claire Villeval. Self control and intertemporal choice: Evidence from glucose and depletion interventions. Working Paper, 2013.

David Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–477, 1997.

David Laibson, Andrea Repetto, and Jeremy Tobacman. A debt puzzle. In Philippe Aghion, Roman Frydma, Joseph Stiglitz, and Michael Woodford, editors, *Knowledge, Information and Expectations in Modern Economics: In Honor of Edmund S. Phelps*, pages 228–266. Princeton University Press, 2003.

David Laibson, Andrea Repetto, and Jeremy Tobacman. Estimating discount functions with consumption choices over the lifecycle. NBER Working Paper No. 13314, 2007.

Emily C. Lawrance. Poverty and the rate of time preference: Evidence from panel data. *Journal of Political Economy*, 99(1):54–77, 1991.

R. Duncan Luce. *Individual Choice Behavior*. John Wiley & Sons, New York, 1959.

Stephan Meier and Charles Sprenger. Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210, 2010.

Stephan Meier and Charles Sprenger. Time discounting predicts creditworthiness. *Psychological Science*, 23:56–58, 2012.

Walter Mischel, Yuichi Shoda, and Monica L. Rodriguez. Delay of gratification in children. *Science*, 244:933–938, 1989.

Ted O'Donoghue and Matthew Rabin. Doing it now or later. *American Economic Review*, 89(1):103–124, 1999.

Matthew Rabin. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292, 2000.

John T. Warner and Saul Pleeter. The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, 91(1):33–53, 2001.

CHAPTER 3

Self Control and Intertemporal Choice: Evidence from Gluscose and Depletion Interventions[75]

Abstract:

Recent economic theories model intertemporal choice as a problem of willpower or self control, *i.e.* of restraining a natural impulse to consume today. We use two interventions that have been shown by psychologists to affect self-control to examine whether this applies to the intertemporal savings decision, as measured in a laboratory elicitation of time preference. Contrary to the predictions of willpower-based models, we find that prior participation in an impulse-controlling activity ("depletion") increases savings. Consistent with those models, sugared-drink consumption raises savings relative to a sugar-free placebo, but the placebo drink also raises savings (relative to no drink) by about the same amount. All these treatment effects are driven by increases in the intertemporal substitution elasticity and are much stronger among subjects with average (as opposed to high) cognitive ability. Overall, we suspect that factors like subjects' attention to the details of the decision are affected by our interventions and are more relevant to the financial decisions we model than are differences in willpower and body-energy budgets.

[75]Co-authored with Peter Kuhn, of the University of California, Santa Barbara, and Marie Claire Villeval of Université of Lyon.

3.1 Introduction

Models of willpower, temptation and self-control are now commonplace in economics. They all aim to capture the visceral push-pull relationship between temptation and prudence that seems a natural way to conceptualize intertemporal choices. Decision makers are modeled as two conflicting selves (Shefrin and Thaler 1988; Bernheim and Rangel 2004; Fudenberg and Levine 2006) or as receiving extra utility when gratification is immediate (Laibson 1997; O'Donoghue and Rabin 1999). Ozdenoren *et al.* (2012) explicitly model willpower as a depletable resource. But how relevant is this modeling approach to fundamental economic behaviors such as borrowing and saving money? By implementing interventions identified by the psychology literature as shifters of willpower, our goal in this paper is to determine whether a willpower-based theoretical paradigm is consistent with across-treatment comparative statics in an abstract borrowing/saving environment.

Psychological research suggests that willpower –the ability to control the self and refrain from impulsive or short-sighted decisions– is negatively affected by prior performance of a task that also requires impulse control. Both dieters who are exposed to the sight of tempting snacks, and those who are asked to suppress their emotional responses while watching an emotional video, subsequently consume more ice cream than dieters engaged in tasks not requiring impulse control (Vohs and Heatherton 2000).[76] Similar effects have been found for performance on a wide variety of tasks.[77] Muraven and Baumeister (2000)

[76]Similar studies of the effects of willpower depletion –manipulated in a variety of ways– on subjects' management of food and alcohol consumption include Kahan *et al.* (2003), Muraven *et al.* (2005) and Baumeister *et al.* (1998).

[77]These include resisting opportunities to cheat the experimenter for financial gain (Mead *et al.* 2009), suppressing stereotypes and prejudice (Gordijn *et al.* 2004; Richeson and Shelton, 2003; Richeson and

argue that these results are consistent with a resource-depletion model of self-control: " controlling one's own behavior requires the expenditure of some inner, limited resource that is depleted afterward." (p.247).

More recently, a number of investigators have argued that blood glucose –which constitutes the body's primary source of energy– is the limited resource that is depleted by acts of self-regulation. In a series of experiments, Gailliot *et al.* (2007, 2009) find that engaging in self-control reduces measured levels of blood glucose, that these induced low glucose levels predict poor performance on a variety of subsequent self-control tasks, and that consumption of a drink sweetened with sucrose (relative to an artificially-sweetened drink) mitigates these poor performance levels.[78] Effects of sucrose consumption have been demonstrated on outcomes including inflicting pain on others, the use of racial stereotypes and slurs, and support for social welfare.[79]

Motivated by the parallels between the psychology experiments and economic theory, this paper considers whether depletion of impulse control and sucrose consumption affect a fundamental aspect of economic behavior: the allocation of income over time. While it might be tempting to view intertemporal allocation decisions as simply a case of

---

Trawalter, 2005; Richeson *et al.*, 2005), restraining aggression (DeWall *et al. 2007*; Stucke and Baumeister, 2006) and impulsive discounting (Hinson *et al.*, 2003). In economics, Bucciol *et al.* (2011) demonstrated that productivity in a task is negatively affected by prior exposure to consumption temptation while Burger, Charness and Lynham (2011) showed that a depleting task –the Stroop (1935) task– improves long-run task completion in a procrastination study.

[78] When sucrose is consumed, glucose is absorbed into the bloodstream at a rate of 30 calories per minute. Metabolization to the brain typically occurs within ten minutes (Donohoe and Benton 1999).

[79] See Aarøe and Bang Petersen (2013) on social welfare and Gailliot and Baumeister (2007) for a review of studies in psychology. The role of nutrition has been little studied in economics. Dotter (2013) has found an impact of breakfast programs in schools on both math and reading scores. While well-being is improved by the consumption of fruits and vegetables (Blanchflower *et al.* 2012), productivity has been shown to increase after individuals were provided chocolates and fruits (Oswald *et al.* 2014). Finally, absorbing glucose seems to impact the mode of reasoning of individuals, and notably increases the likelihood of Bayesian choices over reinforcement heuristic-based choices (Dickinson *et al.* 2013).

resisting an impulse to consume more sooner, it is not at all clear that factors that affect actions like using racial stereotypes or inflicting pain on others will affect financial decisions at all, or in the same way. Adapting the Convex Time Budget technique developed by Andreoni and Sprenger (2012), we allow participants in a laboratory experiment to make a series of choices about payments they will receive sooner or later, across conditions related to willpower depletion and sugar consumption.

In addition to testing a fundamental prediction of willpower-based models of time preferences, our paper makes the following contributions. First, by exposing subjects to a menu of intertemporal choices, we are able to formalize the somewhat imprecise concept of willpower by distinguishing three ways an intervention can affect intertemporal choices: (i) raising the desire to have money sooner rather than later, regardless of whether 'sooner' is today; (ii) raising the desire to have money today, compared to all other options; or (iii) reducing subjects' willingness to sacrifice earlier consumption when its relative cost rises. Loosely, these three effects map into the parameters of a widely used intertemporal utility function –the discount rate, present bias and intertemporal elasticity of substitution– which we structurally estimate for each subject in our sample.[80] Of these, the informal 'willpower' hypothesis probably maps most closely into the present bias parameter (ii). Some alternative cognitive factors, such as, for example, the care or attention subjects are applying to their decisions, may map more closely into the price-sensitivity parameter (iii).

Second, motivated by the literature on cognitive ability and time preferences we

[80]Our experiment is one of very few that is designed explicitly to examine treatment effects on the parameters of structural model. Callen *et al.* (2013) and Carvalho *et al.* (2013) do so with risk preferences and violent trauma and time preferences and savings accounts respectively. This technique can provide a deeper understanding of the treatment effects and allows for more flexible policy analysis.

explore the interaction between our treatment effects and cognitive ability. Previous literature has identified strong correlations between childrens' ability to resist temptation in the famous marshmallow experiment (Mischel, Ebbesen and Raskoff Zeiss 1972) and a variety of cognitive outcomes, including SAT scores (Shoda *et al.*, 1990), IQ (Funder and Block 1989) and college GPA (Kirby *et al.*, 2005).[81] A notable recent finding is that of Benjamin *et al.* (2014), who identify a relationship between cognitive ability and 'behavioral' risk preferences. Both their work and ours contribute to the broader literature studying *which* individuals violate the standard assumptions of economic models, and *when* they do so. In this regard, we generate estimates (from the bottom half of our sample) that are representative of the 50th-90th percentiles of French high school graduates, and we compare the behavior of this group to a very high-ability group (the top half of our sample, which corresponds to the top 10% of high school graduates).

Finally, our study contributes to a recent literature on the impact of temporary manipulations of the decision environment on time preferences. In particular, Ifcher and Zarghamee (2011) find that induced positive affect leads to more patient choices between money received today vs. later.[82] With a within-subject design, Wang and Dvorak (2010) find that sucrose consumption raises patience whereas drinking a sugar-free beverage decreases patience.[83] The key differences between our paper and these two studies are our

[81]Other studies that have linked intelligence with discounting include Frederick (2005), Dohmen *et al.* (2010), Rustichini *et el.* (2012) and Shamosh and Gray's (2008) meta-analysis.

[82]Other manipulations include stress. While Haushofer *et al.* (2013) find no effect of stress induction on patience, Cornelisse *et al.* (2013) find that directly administering the stress hormone, cortisol, *does* make people more present-biased. Using binary choices, none of these studies can estimate the structural model of preferences presented here.

[83]To explain the patience-enhancing effects of sucrose consumption, Wang and Dvorak propose an energy-budget regulation model, in which a high body energy budget makes organisms more future-oriented to facilitate reproduction. In contrast, they argue that the artificial sweetener alerts the body to a caloric crisis,

study of a willpower-depleting intervention and our use of a time-preference task and structural estimation procedure that allow us to distinguish treatment effects on discounting, present bias and price-sensitivity.[84] Together, these aspects of our approach shed light on the mechanisms behind all our treatment effects.

We find that time preferences are sensitive to all our interventions, but we do not find that willpower depletion makes our subjects less patient.[85] Instead, participants who have been exposed to a widely-used willpower-depleting task –the Stroop (1935) test– exhibit *increased* patience in the subsequent time preference elicitation. Consistent with previous studies of impulse control in other contexts, we find that relative to a sugar-free beverage, consumption of a sugared beverage increases patience in our time preference task. Surprisingly, however, this effect is generally smaller in magnitude than the patience-increasing effect of the sugar-free beverage itself (compared to a baseline condition with no beverage), casting doubt on the importance of body-energy budgets relative to situational factors. That said, all our estimated treatment effects are economically significant in magnitude, corresponding to large differences in demand for short-term loans.

Consistent with many studies showing an interaction between cognitive ability and

signaling a low energy budget, which leads to more present-oriented choices.

[84]Wang and Dvorak estimate time preferences by imposing a one-parameter hyperbolic utility function on each subject. A subject's preference parameter (k) is elicited by giving them seven choices between money 'tomorrow' and a future date, with each choice corresponding to indifference at a different level of k. The subject's k is then calculated as the geometric midpoint between the k's of the two options where their choice switches between the early and late option. Since the actual choice options facing subjects before and after beverage consumption were different, this procedure relies heavily on the assumption of a specific one-parameter family of preferences. Also, neither Wang-Dvorak nor Ifcher-Zarghamee gave subjects the choice between consumption at two future dates, so they cannot distinguish discounting from present bias.

[85]Throughout this paper we use 'patience' as convenient shorthand for a tendency to delay the receipt of income, holding other conditions (prices, amount of delay), constant. Since 'patience' is sometimes also used, more specifically, to refer to an absence of present bias in a structural model of choice, we will be explicit whenever we discuss present bias per se.

time preferences, we find that all our treatment effects are stronger in the bottom half of our sample, when ranked by cognitive ability. Subjects with very high cognitive abilities, on the other hand, make decisions that for the most part are unaffected by our attempts to manipulate their willpower. Finally, both our nonstructural and structural estimates indicate that the main utility parameter affected by the treatments (whether depletion, sugared drink or placebo drink) is not the subjective discount rate or present bias parameter, but the intertemporal elasticity of substitution. While subjects in all treatments choose the same level of early income when early income is cheap, treated subjects are much more likely to reduce their early income when its relative price rises.

What processes might explain the tendency of our subjects –excluding those with very high cognitive abilities– to become more price-sensitive in their intertemporal choices when exposed to all our experimental treatments? While we did not set out to test attention-based models of intertemporal choice (such as Koszegi and Szeidl 2013), we note that the effects of all our treatments on subjects' price sensitivity are consistent with attention-enhancing effects.[86] Viewed this way, our estimated treatment effects suggest that exposure to the Stroop test may paradoxically have primed our subjects to pay more attention to the time preference task. Similarly, both the sugared and the placebo beverages may have improved our fasted participants' ability to concentrate on the task, though the effect of the placebo beverage casts doubt on blood glucose *per se* as the primary mechanism for this effect.

[86] One recent study suggesting that attention is a key factor in decision quality is Carroll *et al.* (2009) show that simply manipulating the default savings plan is a powerful tool for increasing individual savings. They also find that forcing an active choice improves choices. In addition, neuroeconomic studies have demonstrated that attention manipulation can improve decision-making in situations involving self-control (Hare *et al.* 2011, Harris *et al.* 2013).

Also consistent with an attention-based explanation, we note that –like all experimental elicitations of time preferences– our procedures elicit preferences for the timing of income, not consumption.[87] Thus, as is well known (see Chabris, Laibson and Schuldt 2008), experimental estimates of intertemporal substitution elasticities which are substantially below infinity imply either that subjects do not have low-cost access to capital markets, or fail to perceive those extra-laboratory options for shifting income over time. Bearing this in mind, estimates of our price-sensitivity parameter may also reflect changes in subjects' awareness of those options. To the extent that cognitively better-endowed persons have a greater tacit understanding of these credit market arbitrage opportunities, this interpretation is also consistent with the fact that our treatment effects largely vanish among higher-scoring subjects.

The remainder of this paper is organized as follows. Section 3.2 details the experimental design. We present our data analysis in Section 3.3 while Section 3.4 discusses our results and concludes.

3.2 The Experiment

3.2.1 Treatments

Our experiment consists of three types of sessions: Baseline, Depletion and Drink. Within each session type, there are five distinct parts, the orders of which change across session type. In a Drink session, the phases are: (1) consumption of drink and entry ques-

[87] Augenblick *et al.* (2012), study the allocation of effort over time. Consistent with the notion that intertemporal arbitrage opportunities are lower for effort than income, they find substantially more evidence of present bias than we do here.

Table 3.1: Experimental Design

	Task				
Treatment	(1)	(2)	(3)	(4)	(5)
Baseline	Entry survey	Rest	Time preference task	Stroop task	Exit survey
Depletion	Entry survey	Rest	Stroop task	Time preference task	Exit survey
Placebo	Sugar-free drink & Entry survey	Rest	Time preference task	Stroop task	Exit survey
Sugar	Sugared drink & Entry survey	Rest	Time preference task	Stroop task	Exit survey

tions, (2) rest to allow any sucrose in the drink to be metabolized into blood glucose, (3) elicitation of time preferences, (4) depletion of self-control in the Stroop test, and (5) an exit survey that includes Frederick's (2005) Cognitive Reflection Test (CRT). The structure of the Baseline sessions is similar to that of Drink sessions, except that no beverage is given. In Depletion sessions, we invert the order between the Stroop test and the elicitation of time preferences. Finally, within the Drink sessions, we have two conditions corresponding to a drink containing sugar or a sugar-substitute. These variations give us four treatments: Baseline, Depletion, Placebo and Sugar. Table 3.1 lays out the progression of the experiment for each treatment.

The comparison between the Depletion treatment and the Baseline allows us to determine whether performing an initial task that requires impulse control affects the decision to defer income in the time preference task. The comparison between the Sugar treatment and the Placebo treatment allows us to study whether the consumption of sugar affects time preferences. Finally, if time preferences react to the consumption and metabolization of sucrose rather than the drink itself, we expect to observe no differences in choices when comparing the Placebo treatment and the Baseline. We discuss each task and drink consumption in more detail below.

3.2.2 Time Preference Elicitation

To elicit time preferences, we implement the Convex Time Budget (CTB) method of Andreoni and Sprenger (2012, henceforth AS). This approach allows us to estimate individual-specific preference parameters.

In every choice, participants received a budget of 16 tokens to allocate between an early payment, c_t, and a late payment, c_{t+k}, with t the early payment date and k the delay between the two dates. Participants made 45 allocation decisions and one of these decisions was randomly selected at the end of the session for actual payment according to the allocation of tokens between the two dates. The 45 budgets combine three early payment dates (t = 0, 5, 15 weeks), three delay lengths (k = 5, 10, 15 weeks) and various price ratios. Thus, there were only seven paydays evenly spaced at five weeks intervals (0, 5, 10, 15, 20, 25, 30 weeks). For each (t, k) combination, participants had to make five decisions involving various interest rates. We defined three rate progressions that were combined with the various early payment dates while the combination of budget progressions and delay lengths were kept constant. The value of a token at the late date, a_{t+k}, was always equal to €1, while the value of the token at the early date, a_t, varied between a minimum of €0.67 and a maximum of €0.99. Allocating all the tokens to the late payment date paid €16; allocating all the tokens to the early payment date paid a minimum of €10.72 and a maximum of €15.84. The progressions were defined in order to offer implied annual interest rates, compounded quarterly, between 4% and 845%. Table A3.1 in the appendix presents all the choice sets.

The presentation of the 45 decisions was very similar to that in AS. A choice screen had nine decision tabs that were displayed successively and corresponded to the nine (t, k) combinations. The order between the nine tabs was randomly and independently determined for each participant to control for order effects. Each decision tab displayed five budget decisions presented in order of increasing gross interest rate. To facilitate decision-making by a better visualization of delays, each decision tab displayed a dynamic calendar highlighting the current date, the early date and the late date in different colors. It also displayed the values of a token at the early date and at the late date, together with the values in Euros of the earnings corresponding to the decisions. A sample decision tab is reproduced in the Appendix. The boxes for entering the allocation decisions were initially blank. As soon as a value was entered either for the early date or the late date, the other box was filled automatically to ensure that the total budget was 16 tokens and the corresponding payoffs in Euro at the two dates were also displayed.

This design allows us to estimate for each individual her discount rate, the curvature of her utility function (through the variations of k and of the gross interest rate), and her present bias (through the variation of t). In the context of our study, it allows us to examine which, if any, of these dimensions is impacted by self-control depletion and sucrose consumption.

3.2.3 Willpower Depletion

We used a Stroop test (Stroop, 1935) to deplete self-control as shown by many studies in social psychology (for a survey of the test, see MacLeod 1991). In a typical Stroop test, individuals have to read the color of ink used to write words independently of the color names of words. In some trials, there is congruence between the color of the word and the color of the ink (the word "yellow" is written in yellow) but in other trials there is no congruence (the word "yellow" is written in red and the correct answer is red). The incongruent stimuli typically require more time and produce more mistakes than the congruent stimuli because the brain automatically decodes the semantic meaning of the word and needs to override its first reaction to identify the color of the ink. Shortcutting the automatic process requires self-control.[88]

In our experiment, the participants' computer screen displayed a series of color words (black, blue, yellow, green and red) successively, and the participants were instructed to indicate, as quickly and accurately as possible, the ink color in which the word was written. The list of possible colors was displayed at the bottom of the screen and the participants had to press the button corresponding to the color of the ink, whether or not that matched the color name of the word (see instructions in appendix). They had to complete congruent and incongruent Stroop trials in random order for 6 minutes. On average they completed 126 trials (S.D. = 11.69). As expected, the time spent on incongruent words was

[88]Note that poor performance in the difficult trials of the Stroop test has been linked to low glucose level (Benton *et al.*, 1994). Study 5 in Gailliot *et al.* (2007) also shows that a lower level of glucose after performing the Stroop test impaired persistence in an additional task. The studies on self-control by Gailliot *et al.* do not only use the Stroop test but they indicate that it is one of the most frequently used measures of self-control.

significantly higher than on the congruent words (two-tailed t-test, $p < 0.001$).

3.2.4 Drink Consumption

Following Gailliot *et al.* (2007), participants in each Drink session were given 14 ounces (40 centiliters) of a soft drink sweetened either with sugar or with a sugar substitute. Both types of drinks had the same appearance. The sugared drink contained 158 kilocalories and the placebo drink contained 10.[89] We used a double blind procedure to administer the drinks: neither the participants nor the experimenters were aware of the sugar content of the beverage.

After being invited to drink the beverage, participants could rest in silence and read magazines that we distributed during 10 minutes in order to allow the sucrose to be metabolized into glucose. Three minutes before the end of this period, participants had to assess the beverage and to report their usual consumption of soft drinks.[90] In the Baseline and the Depletion treatments, the same rest period of 10 minutes was implemented.

[89] Specifically, the drinks were Fanta "Citron frappe" and Fanta Zero "Citron frappe". They were dispensed in glasses (not the original container) and appear identical (see Figure A3.1 in the appendix). Neither contains caffeine, though both contain ascorbic acid (vitamin C).

[90] The questions were: 1) Please rate your enjoyment of the beverage you just consumed, between 1 and 10. 2) How many calories do you think the beverage contained? 3) How often do you drink soft drinks (Coke, Pepsi, lemonade, ...): every day / every week / once or twice a month or less / less than twice a month? Although participants in the Placebo condition assessed the beverage less positively (mean = 4.55, S.D. = 2.77) than those in the Sugar condition (mean = 5.57, S.D. = 2.58) (two-tailed Mann-Whitney test, p = 0.097), they did not realize that they received a placebo. Indeed, they predicted the same number of calories contained in the beverage (mean = 124.16, S.D. = 86.26) than the participants placed in the Sugar condition (mean = 140.41, S.D. = 98.26) (p = 0.497).

3.2.5 Procedures

The experiment was computerized, using the REGATE-NG software. It consisted of 8 sessions conducted at the laboratory of the GATE (Groupe d'Analyse et de Théorie Economique) institute in Lyon, France. Undergraduate students from the local engineering and business schools were invited via the ORSEE software (Greiner 2004). Between 17 and 20 participants took part in each session, for a total of 149 participants. Two sessions of the Baseline treatment were implemented with a total of 34 participants; two sessions of the Depletion treatment were implemented involving 40 participants; and four Drink sessions were implemented with 75 participants (37 in the sugar condition and 38 in the placebo condition).

The invitation message addressed to the participants of all treatments indicated that they may possibly have to drink a beverage containing sugar during the session and that individuals suffering or thinking that they may suffer from a pathology linked to blood glucose regulation (like diabetes) should abstain from participating. After signing up, all the participants in all the treatments were instructed not to drink or eat at least three hours prior to the beginning of the session in order to stabilize blood glucose levels. Upon arrival we recorded the time of their last intake. Since chronobiology may influence economic decision-making, all the sessions were run at noon, when the level of blood glucose is low.[91]

Upon arrival, the participants had to sign a consent form reminding them that they should not participate if they suffer from a disease related to failure of blood sugar regula-

[91] We did not measure baseline blood glucose levels, which would have required taking blood samples.

tion. Then participants randomly drew a tag from a bag assigning them to a terminal. The instructions for each segment were distributed and read aloud by the experimenter after the completion of the prior segment (see appendix).

The elicitation of time preferences requires very strict procedural rules. To participate in the experiment, the students were required to own a personal bank account and were informed by the invitation message that they would be paid by a wire transfer to their bank account; a bank statement was required.[92] During the session, instructions informed the participants that a show-up fee of €5 would be wired to their bank account in addition to their other payoffs at two different dates, regardless of their decisions: half of the show-up fee amount would be paid at the early date and the other half at the late date indicated by the decision randomly selected at the end of the session for payment. The show-up fee had no differential influence on the 45 allocation decisions. Participants were also informed that the dates mentioned on the decision screens were the dates at which the wire transfers would be ordered by the finance department.[93] To maximize the confidence of the participants about the payment of their earnings, they received a document stating that the bank transfer would be ordered by the National Center for Scientific Research (CNRS).[94] In addition, the document mentioned the name, email address and phone number of the

[92]We cannot rule out that the information given in the message (payment wired to the bank account and possibility of having to drink a beverage) has led to a self-selection of participants. However, the sessions were booked as quickly as usual. In addition, we asked 44 students participating in another experiment with standard cash payment whether they owned a personal bank account; all of them answered positively. Moreover, there is no reason to believe that the two criteria for participating were correlated. Finally, the message did not mention that the payment could be made at two different dates.

[93]The administration committed to respect exactly the dates of the transfers and sent us a feedback after each payment. We believe the transaction costs associated with this payment methodology are lower than the typical approach used in this type of experiment, which relies on personal checks or vouchers.

[94]In France, CNRS is a well-known science and technology public agency. It employs 25,000 people and it operates through 1,235 research institutes. Students are aware that the GATE institute is operated by both the CNRS and the University of Lyon.

professor in charge of the experiment who could be contacted in case of any problem with the payment.

At the end of each session, participants received a feedback on the decision randomly selected for payment, indicating their payoffs and the dates of the two wire transfers for this decision. Then, they had to complete an exit survey which included questions about their demographics and average mark on the final high school exam (Baccalauréat). Sessions lasted 60 minutes and participants averaged earnings of 20.43 ($26.62, with a standard deviation of €0.97 ($1.26), including the show up fee.

3.3 Results

We present our results in four sections. The first section establishes a number of basic patterns in a pooled sample of all treatments, to provide context for the study of treatment effects. The second and third sections are nonparametric and structural approaches to analyzing the treatment effects, respectively. The final section presents some robustness checks. Since -as noted– one of our central questions is how the impact of treatment manipulation on patience is mediated by the subjects' cognitive ability, we present all of our experimental results separately according to our subjects' reported achievement on the French Baccalauréat exam.[95] We divide our participants in half relative to the median score in our sample, which was 16; this is also an important cutoff in the distribution of scores for student achievement.[96] Importantly, because only 9% of French Baccalauréat recipients

[95] The French Baccalauréat exam is taken at the end of high school (*lycée*). In 2012, slightly over three quarters of French youth had passed the Baccalauréat.

[96] We also consider fuller distributional effects and effects based on lab-elicited CRT score.

earned a score of 16 or higher (our participants are drawn from selective universities), we refer to our two groups as “high score” and “lower score” respectively. Our high-scoring subjects clearly represent an elite level (about the top decile) of achievement among French high school graduates, while our lower-scoring group roughly represents the 50th through 90th percentiles. Thus the results for our lower-scoring group are more representative of a typical high school graduate in France, and we focus much of our discussion on that group.[97]

3.3.1 Overall Features of Behavior

We start by presenting two foundational results that verify aspects of our model and design, plus some simple descriptive statistics for the pooled sample across all treatments. The first result is that subjects’ aggregate demand curves in the experiment satisfy two general predictions of utility-maximizing intertemporal behavior.

Result 1 - Consistent with predictions for agents who discount the future and have some preference curvature, mean demand for early income exceeds half the 16-token endowment at interest rates near zero, then declines monotonically with the price of early income. This behavior characterizes both high- and lower-score participants.

A simple but general model of choice between early and late tokens for any combi-

[97] The results of the Cognitive Reflection Test (CRT) performed at the end of the sessions are highly correlated with the Baccalauréat score, and we can replicate our main results using this measure of cognitive ability as well. However since subjects’ CRT results could be affected by our treatments, we focus on the Baccalauréat-score based results.

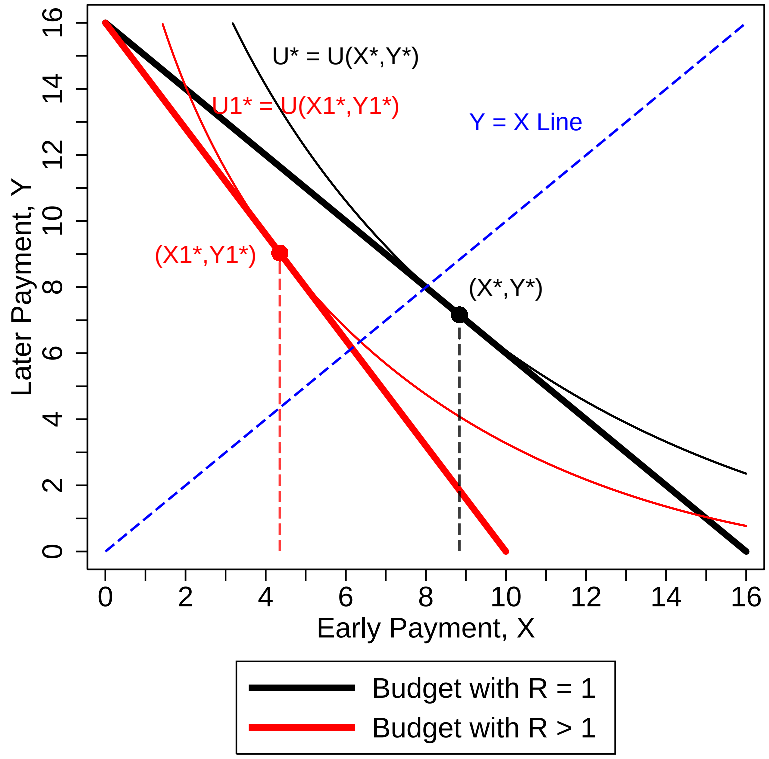


Figure 3.1: Predicted Behavior

nation of early payment date (t) and delay (k) supposes that subjects solve

$$\max_{X,Y} \quad U(X) + \lambda U(Y), \quad \text{subject to} \quad RX + Y \leq M \tag{1}$$

where X is experimental income received in the early period, Y is experimental income received in the later period, $U' > 0$, $U'' < 0$, R is the price of sooner income, and M is the endowment. In (1), $\lambda < 1$ can depend on both t and k to incorporate both discounting and present bias, but is fixed within any (t, k) cell. R, on the other hand, varies within a (t, k) cell as we experimentally manipulate the implied interest rate. For this model of preferences, Figure 3.1 illustrates (a) that subjects should consume more than half their endowment in the early period ($X > 8$) when $R = 1$ because $\lambda < 1$, and that X should fall monotonically as R rises because income and substitution effects reinforce each other when the endowment is all in the later period, as is the case in our experiment.[98]

Figure 3.2 plots the data-generated demand curves for the early payment (X), separately by score and pooled across all treatments. With the exception of the shortest delay length and latest start date for both groups, the demand curves all start at above eight units of X at levels of R closest to one, then fall monotonically as R rises.[99] The success of these basic predictions suggests that our participants' choices are informative for the preferences we wish to study.

Result 2 - There is evidence of small but significant present bias in our data, among both high- and lower-test score participants.

Participants receive the first of their two payments either on the day of the experiment, 5 weeks after the experiment or 15 weeks after the experiment. To test formally for present bias we regress early payments on dummy variables for $t = 5$ and $t = 15$ as well as

[98]Alert readers will note that equation (1) models demand for early versus late experimental *payments* in the same way economists typically model intertemporal *consumption* choices. Of course, if subjects choose total consumption according to (1) but have access to perfect capital markets, their demand for experimental payments will consist of corner solutions (i.e. either $X = 0$ or $Y = 0$) that maximize the market value of experimental payments. Effectively, subjects would behave as if the U function had little or no curvature. We test this idea formally in Section 3.3.3 and argue that it may shed some light on the possible mechanisms behind our estimated treatment effects.

[99]Because we do not observe choices from a zero-interest budget and Figure 3.2 indicates substantial nonlinearity in the demand curves, we used our structural model to estimate choices at $R = 1$ to further test the prediction about income levels when $R = 1$. We find strong support, for all combinations of delay length and whether the early payment occurs immediately. The minimum predicted zero-interest demand is €9.32 (S.E. = 0.25).

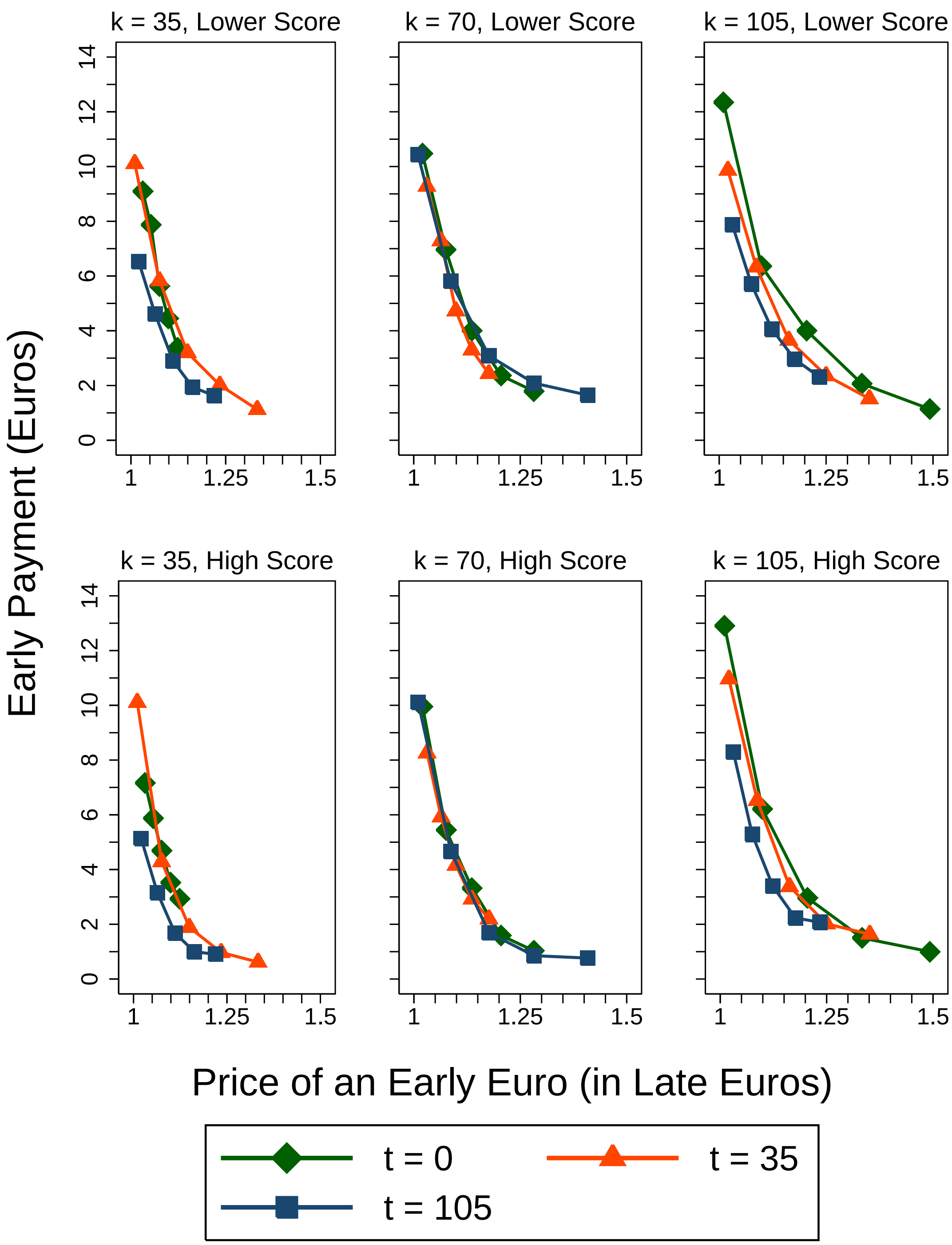


Figure 3.2: Demand Functions by Date of Early Payment, t, All Treatments

Table 3.2: Effect of Start Date, t, on Early Payment Demand

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
Constant ($t = 0$, $R = 1$)	8.256	8.679	7.840
	(0.437)	(0.608)	(0.627)
$1(t = 5$ weeks)	-0.521***	-0.678**	-0.367
	(0.192)	(0.264)	(0.278)
$1(t = 15$ weeks)	-1.324***	-1.308***	-1.340***
	(0.286)	(0.409)	(0.403)
Normalized Price Ratio ($R - 1$)	-21.365***	-21.535***	-21.197***
	(1.197)	(1.723)	(1.675)
Clusters	149	74	75
Observations	6705	3330	3375

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster.

the price ratio while clustering standard errors at the individual level.[100] Table 3.2 presents the results of these regressions. If the date of first payment is immediate rather than 5 or 15 weeks in the future, lower-score subjects borrow significantly more of their endowment. High-score subjects do the same for only the 15 week delay.

Finally, we note that there are only small and statistically insignificant differences between the early payment choices of high- and lower-Baccalauréat-score participants in our overall sample, which combines all treatments. Specifically, lower-score participants select a slightly higher overall level of early payment, and display slightly more present bias (which may be taken as a proxy of impulsivity), but neither gap is significant at conventional levels.[101] As the next section shows, however, this aggregate result obscures

[100] A regression approach is necessary because price ratios are not exactly balanced across the t dimension.

[101] Averaged across all choices, lower-score subjects allocate about €0.70 more experimental income (S.E. = 0.54, clustered by individual) to the earlier payment date than high-score participants. This difference is

sizeable differences in the effects of treatment on the behavior of high- versus lower-score participants.

3.3.2 Simple Estimates of Treatment Effects

Our first look at the effects of the various treatments is non-parametric. Figure 3.3 presents the mean demand for early payments across the Baseline, Depletion, Placebo and Sugar treatments by Baccalauréat score. Since these comparisons are between individuals, the treatments are balanced with respect to prices, delays and start dates.

Result 3 - For the lower test score sample, depletion, a sugared drink and a non-sugared drink all reduce the demand for early payment. All of these treatment effects are absent among participants with very high test scores.

Relative to the baseline condition, all three treatments significantly reduce demand for early income amongst lower-score participants. The strongest effect is for the sugar treatment (average demand for early payment reduced by 51% relative to baseline), with reductions of 38% and 27% for the depletion and placebo treatments respectively. The difference between the the sugar and placebo effects is significant (p = 0.056), suggesting some biologically-based effects of blood glucose.[102] The magnitude of this additional sugar

not significant. We add interaction terms between the dummy variables for t = 5 and t = 15 and high-score as well as a high-score level effect into the present bias regressions from Table 3.2. The gap between early demand when t = 0 versus t = 5 is about €0.31 smaller for high-score participants, but this difference is not significant (S.E. = 0.38). The signs and significances of the non-interacted dummies are unaffected.

[102]We use participants' estimates of the calories their beverage contained in order to ascertain whether this difference is due to psychology or physiology. Amongst lower-score subjects, there is no evidence that the magnitude of the Sugar-Placebo gap is affected by the beliefs about the drink or that beliefs themselves

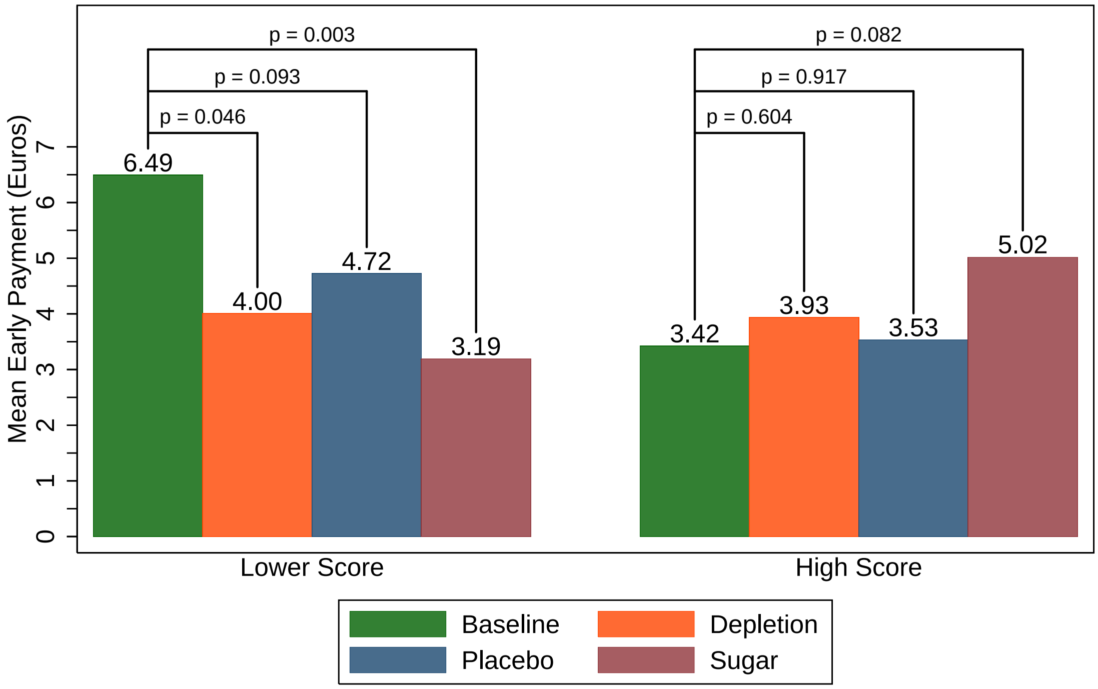


Figure 3.3: Mean Demand by Treatment

p-values are generated from regressions of the chosen early payment on treatment status with standard errors clustered at the individual level. The regression is run separately for lower- and high-score subjects. Each individuals makes 45 decisions, leaving us with a sample size of 3330 (74 clusters) in the lower-score group and 3365 (75 clusters) in the high-score group. An approach that collapses the data to individual-level means yields similar results.

effect (32% relative to the Placebo) is roughly the same size as the initial beverage effect, suggesting a modest role for body energy budgets relative to the other situational factors. Turning to the high-scoring subjects, only the sugar treatment affects the demand for early income significantly (and *positively*) of the high-score subjects (average demand increased by 47%). This effect is not precise, however. It does not significantly differ from the placebo effect at conventional levels (p = 0.143). Tobit regressions that specify censoring points at the corner solutions obtain results that are qualitatively identical.

Another noteworthy aspect of Figure 3.3 is that high- and lower-score subjects differ

generate differences in demand.

substantially in their Baseline choices; the difference of €3.07 between the groups' early payment demand in the Baseline is highly significant (p = 0.011). Recalling that there was no overall difference between high- and lower-score participants, this suggests that our three interventions have the effect of narrowing the difference in choices between high- and lower-score participants by reducing lower-scoring participants' demand for early payments. The next result probes the sources of these effects further.

Result 4 - The negative effect of all three treatments on lower-score participants' demand for early payments is strongest in cases where the price of early income is high.

Figures A3.2 and A3.3 in appendix plot the demand curves for early payments for each (t, k) pair for lower-score and high-score participants, respectively. The lower-score subjects exhibit a similar level of demand across all treatments at low price levels. As the price of early income rises, early payments decline more rapidly in the Depletion, Sugar, and Placebo treatments than in the Baseline. The high-score participants show a similar level of demand to the lower-score participants at low prices, but demand is highly price-sensitive in all four treatments. Thus, the treatments make the lower-score participants more price-sensitive, and thus more similar to the high-score participants' behavior.

To determine the statistical significance of the above effects, we define three price levels based on the relative value of early tokens. When early tokens are worth €0.90 or more we say the price is low, when they are worth between €0.80 and €0.90, we say the

Table 3.3: Treatment Effect on Early Payment Demand by Price Level

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
Constant (Low price, Baseline)	7.976	8.809	6.449
	(0.803)	(1.036)	(1.146)
Low price X Depletion	-1.778	-2.585	-0.276
	(1.093)	(1.590)	(1.460)
Low price X Placebo	-0.865	-0.944	-0.492
	(1.074)	(1.316)	(1.703)
Low price X Sugar	-0.530	-2.569	1.444
	(1.073)	(1.567)	(1.433)
Medium price	-4.423***	-3.848***	-5.477***
	(0.534)	(0.645)	(0.876)
Medium price X Depletion	-1.287	-2.559**	1.171
	(1.862)	(1.171)	(0.844)
Medium price X Placebo	-1.416*	-2.578**	0.789
	(0.793)	(1.040)	(0.754)
Medium price X Sugar	-1.194	-4.321***	2.024**
	(0.803)	(0.943)	(0.657)
High price	-5.764***	-5.550***	-6.157***
	(0.739)	(0.831)	(1.038)
High price X Depletion	-0.806	-2.019**	1.264*
	(0.738)	(1.010)	(0.653)
High price X Placebo	-1.393**	-2.271**	0.268
	(0.686)	(0.967)	(0.454)
High price X Sugar	-1.208*	-3.020***	0.995**
	(0.686)	(0.883)	(0.441)
Clusters	149	74	75
Observations	6705	3330	3375

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster.

price is medium and when they are worth €0.80 or less, we say the price is high.[103] Table

[103]Note that this definition focuses on the most salient aspect of the price presented to the participants: the changing value of an early token within a particular choice screen (t, k combination). Thus, the ranking is

3.3 presents OLS regressions of early payment demand on the treatment dummy variables split by price level. At medium and high prices all three treatments have significant effects for the lower-score group and the magnitude of the sugar effect is larger at high as opposed to low prices. The sugar effect is significantly greater than the placebo effect in the medium price condition (p = 0.003) and borderline significantly greater in the high price condition (p = 0.105). At both medium and high prices, the effect of the placebo beverage (relative to the baseline) is larger in magnitude than the difference between the sugared- and sugar-free drink.[104] Column (3) suggests an elasticity-reducing effect of the Sugar treatment on the high-score subjects, but the effects are not statistically different from the Placebo effects in either medium or high price condition (p = 0.132 and p = 0.218 respectively).

In sum, our nonparametric analysis shows that all three treatments (Depletion, Placebo and Sugar) reduce early demand among subjects with lower Baccalauréat test scores, who are more representative of the educated French population than our high-score sample. This apparent increase in 'patience' occurs only when the price of early income is high, so the treatments effectively make lower-score subjects more price-sensitive and therefore their overall behavior more similar to our 'elite' sample.

different than one based on annualized interest rate.

[104]In the medium-price case, the sugar-free drink reduces early consumption by €2.578, while the additional effect of adding sugar to the drink is a reduction of (€4.321 - €2.578 =) €1.743. In the high price case, these two effects are €2.271 and €0.749 respectively. The *p*-values associated with these comparisions are 0.546 and 0.219 respectively.

3.3.3 Treatment Effects in a Structural Model of Time Preferences

To measure whether the treatments affected different aspects of participants' preferences,[105] we now estimate a simple structural model of intertemporal preferences in which the treatments can affect each one of the fundamental utility parameters (specifically, their discount rate, present bias and intertemporal substitution parameters). One primary advantage of the CTB method is that it allows for the precise estimation of the parameters of structural models of intertemporal choice, even on the individual level. We will consider two types of structural treatment effects: aggregate and individual. Aggregate effects compare one treatment-specific parameter estimate to another and individual effects compare the set of individual-specific parameter estimates within one treatment to those from another. The two approaches yield similar results. As in section 3.3.2, splitting the sample by test score is essential for understanding the treatment effects.

We first provide a characterization of an individual's decision problem. Consider individual i making decision j. Continue to denote X as the number of tokens received at the earlier date and Y the number at the later date. Individual i is assumed to have power income utility (with exponent α) that is additively separable across time periods in a β-δ form (Laibson 1997; O'Donoghue and Rabin 1999). Choice j is characterized by the price of sooner income, R, a delay between the two payment dates, k, and an indicator for whether or not the sooner date is today, T (equal to 1 if t = 0, and 0 otherwise). As in

[105]For example, while reduced utility curvature (higher α) is associated with higher price-sensitivity, it should also increase the response to k (the gap between the payment dates). In general, because the demand functions implied by most theoretically interesting demand functions are nonlinear, the predicted marginal effects of each parameter depend on the levels of all the others, making simple regression tests only roughly informative about the effects of treatments on preference parameters.

equation (1), M is the total number of tokens available. We suppose that subjects optimize in the following way:[106]

$$(X_{ij}, Y_{ij}) = \underset{X,Y}{\arg\max} \quad X^{\alpha} + \beta^{T_j}\delta^{k_j}Y^{\alpha} \quad \text{subject to} \quad R_jX + Y \leq M. \tag{2}$$

To identify preferences, we follow the approach of AS by applying non-linear least squares (NLS) to the demand function for sooner tokens, derived directly from equation 2. This approach yields the structural regression equation

$$X_{ij} = \frac{M(\beta_j^T \delta_j^k R_j)^{\frac{1}{\alpha-1}}}{1 + R_j(\beta_j^T \delta_j^k R_j)^{\frac{1}{\alpha 0-1}}} + \epsilon_{ij}. \tag{3}$$

To analyze and test treatment effects, we replace α with

$$\alpha_1 + \alpha_2 D_i + \alpha_3 P_i + \alpha_4 S_i, \tag{4}$$

where D, P and S are treatment indicator variables, and make similar substitutions for β and δ. Instead of presenting results on δ itself, we use $r = \delta^{-365} - 1$, the yearly discount rate equivalent, for ease of interpretation.

Setting out the structural form in (2)-(4) allows us to be more precise about how our manipulations of the cognitive and physiological environments affect subjects' intertemporal choices than the more generic notions of 'impatience' or 'impulsivity'. For example, if

[106]Note that equation (1) implies that the set of available allocations is convex: that the tokens can be infinitely divided. While we offer subjects 17 possible allocations along the budget frontier rather than an infinite number, we argue that this is a suitable approximation to convexity. Andreoni, Kuhn and Sprenger (2013) perform a similar exercise with 6 allocations and find no evidence of bias due to discretization.

a treatment raises r, it should increase subjects' demand for early rewards relative to late rewards regardless of the amount of delay between the two payment dates, and regardless of whether the early period corresponds to the date of the experiment or a future date. If a treatment lowers β (the present bias parameter) below 1, it increases subjects' attraction only to rewards that are received on the date of the experiment; high levels of present bias (low values of β) generate temporal inconsistencies in choices that may correspond to psychological notions of a failure of willpower (*i.e.* a greater impulsiveness). Finally, if treatments increase α, they make subjects more responsive to the costs of early income, which under some conditions (i.e. access to capital markets) might also be interpreted as an 'improvement' in the effectiveness of subjects' decisions. All three notions are conflated in the more amorphous notion of willpower that is often used to interpret experimental results on the effects of willpower depletion.

We first estimated equation (3) without treatment effects, following our modification of the CTB technique introduced by AS in the calibration of prices. Our estimate of the aggregate yearly discount rate is 21.8% for lower-score types (S.E. = 5.9%) and 21.0% for high-score types (S.E. = 4.1%).[107] Our estimate of the β parameter is 0.976 (S.E. = 0.008) for lower-score and 0.988 (S.E. = 0.007) for high-score, with both values significantly less than 1 (p = 0.005 and p = 0.086, respectively). Thus, in contrast to AS who estimate β = 1.007 (S.E. = 0.006), we find evidence of present bias in the β-δ form.[108] Lastly, we estimate a lower degree of curvature –α = 0.922 (S.E. = 0.008) for lower-score and 0.942

[107] The corresponding specification from AS (Table 3.2, column (3)) estimates a rate of 37.7% with a standard error of 8.7%. Because our max time horizon is slightly longer, we would expect a slightly lower estimate of the rate if individuals display some insensitivity to the exactness of dates far in the future.

[108] While this magnitude of present bias over pure allocations of money is not economically meaningful in our experiment, a 3% distortion of preferences could be very important for major financial decisions.

(S.E. = 0.005) for high-score individuals– as opposed to 0.897 (S.E. = 0.009) in AS.

Result 5 - The treatment effects on the structural parameters are concentrated on α, the utility function curvature parameter. The magnitudes are economically significant at interest rates that correspond to predatory credit instruments.

Table 3.4 presents estimates of treatment effects on the parameters of a common utility function, shared by all individuals in each estimation sample. The treatment effects only show up as significant for utility curvature. Both Drink treatments significantly decrease lower-score curvature (Baseline α = 0.860, Placebo α = 0.946, Sugar α = 0.965), and the estimates are precise enough to conclude the effect is significantly larger for the sugared drink (p = 0.043). The marginal effect of sugar of α (0.105 - 0.087 = 0.018) is much smaller than the effect of the placebo beverage (0.087). The Depletion treatment has a weaker effect (Depletion α = 0.917). Additionally, the joint hypothesis that the α and β effects are zero is rejected (p = 0.088). The two Drink treatments have significant joint effects as well; the joint effects on all three parameters are jointly different from zero in the Placebo treatment (p = 0.021) and in the Sugar treatment (p = 0.001) for the lower-score sample. While the high-score curvature increase is significant only for the Sugar treatment (Baseline α = 0.961, Sugar α = 0.931), this effect is not significantly different from the effect of the Placebo treatment (p = 0.167).

Table 3.4: Treatment Effects on Aggregate Utility Parameter Estimates

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
α **(Utility Curvature)**			
Constant (Baseline)	0.904	0.860	0.961
	(0.015)	(0.027)	(0.007)
Depletion Effect	0.028	0.058*	-0.016
	(0.018)	(0.031)	(0.013)
Placebo Effect	0.042**	0.087***	-0.014
	(0.016)	(0.028)	(0.013)
Sugar Effect	0.036**	0.105***	-0.030***
	(0.016)	(0.028)	(0.011)
β **(Present Bias)**			
Constant (Baseline)	0.979	0.949	1.002
	(0.016)	(0.026)	(0.013)
Depletion Effect	0.006	0.045	-0.023
	(0.018)	(0.027)	(0.018)
Placebo Effect	0.004	0.031	-0.014
	(0.018)	(0.029)	(0.018)
Sugar Effect	0.004	0.025	-0.016
	(0.019)	(0.029)	(0.018)
r **(Annual Discount Rate)**			
Constant (Baseline)	0.268	0.357	0.210
	(0.106)	(0.225)	(0.068)
Depletion Effect	-0.140	-0.267	-0.057
	(0.124)	(0.256)	(0.097)
Placebo Effect	-0.046	-0.076	-0.076
	(0.121)	(0.236)	(0.114)
Sugar Effect	-0.016	-0.219	0.109
	(0.122)	(0.237)	(0.107)
Clusters	149	74	75
Observations	6705	3330	3375

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster.

The fact that the treatment effects operate through utility curvature is consistent with Result 4: they make the lower-score individuals more price sensitive. In the limiting case where the utility function has no curvature, optimal choices move from one corner to the other as prices change. The less curvature the function has, the closer we are to this case, and the more responsive individuals will be. The more curvature the function has, the more we should observe choices that don't respond fully to extreme prices. To illustrate this, consider subjects from our experiment making a decision about taking a 2-week payday loan against a €1000 paycheck that comes with a 15% charge (APR = 390%). Roughly, the optimal loan for a lower-score, Baseline treatment individual is €310, which results in a €60 charge. Holding the discount and present-bias factors constant and switching to the Depletion curvature estimate reduces the loan to €220 (charge of €40), the Placebo curvature estimate to €140 (charge of €20) and the Sugar curvature estimate to €60 (charge of €10).

Turning now to our method that allows each subject to have his/her own set of utility parameters, (α, β and r), we make a couple of adaptations that are dictated by the estimation results. First, we drop 21 individuals who lack enough choice variation for the successful estimation of the parameters. Second, because using the NLS technique with only 45 observations per subject delivers some extreme outlying estimates, we trim the sample at the 5th and 95th percentiles of the distribution of all three parameter estimates. This excludes 24 more subjects, leaving a sample of 104. Of the 45 excluded subjects, 28 are from the lower-score sample and 17 are from the high-score sample.

Table 3.5 reports estimates of treatment effects on the individual-specific parameters

Table 3.5: Treatment Effects on Median Individual Utility Parameter Estimates

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
α **(Utility Curvature)**			
Constant (Baseline)	0.958	0.940	0.974
	(0.009)	(0.011)	(0.008)
Depletion Effect	0.016	0.023*	0.008
	(0.010)	(0.013)	(0.014)
Placebo Effect	0.012	0.027*	0.005
	(0.012)	(0.015)	(0.019)
Sugar Effect	0.010	0.039***	-0.011
	(0.011)	(0.013)	(0.011)
β **(Present Bias)**			
Constant (Baseline)	0.979	0.949	1.013
	(0.026)	(0.028)	(0.014)
Depletion Effect	-0.002	0.041	-0.035**
	(0.027)	(0.030)	(0.016)
Placebo Effect	0.013	0.047	-0.023
	(0.026)	(0.030)	(0.019)
Sugar Effect	-0.001	0.018	-0.035*
	(0.029)	(0.037)	(0.019)
r **(Annual Discount Rate)**			
Constant (Baseline)	0.323	0.490	0.323
	(0.104)	(0.250)	(0.073)
Depletion Effect	0.048	0.017	-0.081
	(0.143)	(0.285)	(0.140)
Placebo Effect	0.105	-0.040	-0.034
	(0.148)	(0.272)	(0.192)
Sugar Effect	-0.000	-0.311	0.109
	(0.130)	(0.273)	(0.149)
Observations	104	46	58

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Standard Errors in parentheses. Parameter estimate distributions trimmed at the 5th and 95th percentiles.

using quantile regressions at the median value of the estimate distribution. Specifically, for each of the three parameters, we estimated a median regression on 104 observations in which the participant's parameter estimate was the dependent variable and the three treatment indicators were the only regressors. Standard errors for these estimates are obtained via bootstrap. The estimated individual effects are largely consistent with the aggregate effects. Both drinks significantly decrease curvature in the lower-score sample, whereas the depletion effect on curvature is weaker but still marginally significant. One puzzle is that the treatments appear to have present-bias inducing effects for the high-score group in this specification.

All three treatments increase the amount of deferred income for the lower-score individuals by reducing utility curvature such that budgets featuring above-market interest rates generate large differences in allocations versus the Baseline. There exists some evidence that the Sugar treatment had stronger effects than the Placebo treatment.

3.3.4 Robustness

While our structural estimation procedure uses both interior and 'corner' choices to identify the utility parameters, and while the procedure is compatible with any finite level of intertemporal substitutability, a possible concern is that the method breaks down in the limiting case of infinite substitutability across time periods, where all choices are predicted to be at corners.[109] Since our structural estimates suggest a high degree of substitutability,

[109]This is a consequence of assuming that individuals evaluate the utility of lab earnings as prospects independent of background consumption. One approach to this issue would be to incorporate background payments into the structural estimation. Since background consumption is independent of treatments this would have little effect on the results.

Table 3.6: Corner Choice by Treatment

Treatment	All Subjects (1)	Lower-Score (2)	High-Score (3)
Baseline	24% Sooner Corner 47% Later Corner 29% Interior	29% Sooner Corner 34% Later Corner 37% Interior	16% Sooner Corner 71% Later Corner 13% Interior
Depletion	18% Sooner Corner 62% Later Corner 20% Interior	15% Sooner Corner 56% Later Corner 29% Interior	21% Sooner Corner 68% Later Corner 11% Interior
Placebo	19% Sooner Corner 58% Later Corner 23% Interior	21% Sooner Corner 52% Later Corner 27% Interior	15% Sooner Corner 67% Later Corner 18% Interior
Sugar	18% Sooner Corner 53% Later Corner 29% Interior	13% Sooner Corner 65% Later Corner 22% Interior	20% Sooner Corner 49% Later Corner 31% Interior

and since a substantial share of our subjects' choices are, in fact, at corners, we also studied treatment effects on the frequency and type of corner solutions in two less parametric ways. The first of these, in Table 3.6, shows the frequency of interior solutions and the two types of corner solution, by treatment.

While the overall share of corner solutions in Table 3.6 is high at 75% (consistent with AS and with Andreoni *et al.* 2013), column (2) also clearly shows that all treatments reduce interior choice frequency among our lower Baccalauréat score participants. The especially pronounced increase in later corner choices for this group is related to the economically, but not statistically significant changes to the discount rate induced by the treatments.

Our second approach was to estimate treatment effects in a multinomial logit specification with three choice options: 1) sooner corner, 2) interior and 3) later corner. Results are found in appendix Table A3.4. Reassuringly, in the lower-score sample, the probability of choosing the sooner corner is significantly lower in the Depletion and Sugar treatments than the Baseline, and the probability of choosing the later corner is significantly greater in the Depletion, Sugar and Placebo treatments.

If time preferences are indeed dependent on physiological conditions, it would be encouraging if our treatment effects were moderated by the condition in which individual subjects entered the lab. While subjects were asked not to eat or drink for at least three hours prior to the experiment, our survey indicated that there was substantial variation in the degree of adherence to this request. Almost 19% of individuals report they had not eaten since the day before the experiment and around 7% had eaten within the three hour window prior to the experiment. We expect that subjects should have been more susceptible to the interventions the longer they went without eating. Table 3.7 presents treatment effect regressions on demand for early payment with interactions between the Depletion, Placebo and Sugar variables with the number of hours since last meal.

Consistent with our baseline results, we find no significant meal-time correlations for the high-scoring subjects; this group's decisions are also unaffected by the amount of elapsed time since their last meal. Lower-scoring subjects, on the other hand, become less patient as the time since their last meal increases; this behavior is consistent with Briers *et al.*'s (2006) and Danziger *et al.*'s (2011) evidence.[110] Also, as predicted, lower-scoring subjects' sensitivity to all three of our interventions increases with elapsed time since their

[110]Briers *et al.* found that the desire for caloric resources increases the desire for money. Looking at decisions made by an Israeli parole board, Danziger *et al.* found that parole was much more likely to be granted early in the day than later in the day, conditional on crime, sentence and ethnicity. Since a judge's reputation is harmed more by inappropriately granting, as opposed to inappropriately refusing parole, fatigued judges 'take the easy way out' relative to rested judges. Following the board's midmorning snack, there was a substantial spike in the percentage of prisoners who were granted parole.

Table 3.7: Treatment Effect on Early Payment Demand with Meal Time Controls

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
Constant (Baseline, just ate)	2.838	2.360	3.735
	(1.248)	(1.287)	(1.509)
Depletion Effect	1.357	1.281	1.401
	(1.645)	(2.069)	(2.079)
Placebo Effect	0.080	1.419	-2.326
	(1.524)	(1.630)	(2.159)
Sugar Effect	2.591*	1.102	2.090
	(1.526)	(1.700)	(1.844)
Time since last meal (hours)	0.434**	0.730***	-0.049
	(0.218)	(0.155)	(0.223)
Time X Depletion	-0.474*	-0.677***	-0.217
	(0.261)	(0.241)	(0.301)
Time X Placebo	-0.183	-0.544**	0.421
	(0.246)	(0.214)	(0.289)
Time X Sugar	-0.608**	-0.773***	-0.122
	(0.250)	(0.232)	(0.270)
Clusters	149	74	75
Observations	6705	3330	3375

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster.

last meal.[111] While this may not be surprising for the drink treatments, it is perhaps noteworthy that the Stroop test also has a larger patience-enhancing effect on hungry than on recently-nourished subjects. This finding reinforces our suggestion that engaging in a cognitively demanding task that requires resisting one's immediate impulse can (at least temporarily) improve a vulnerable subject's ability to focus on subsequent economic decisions.

[111] Note that the uninteracted treatment effects no longer enter as significant because they are estimates specific to the intercept where the time since last meal is zero.

Table 3.8: Treatment Effect on Early Payment Demand by CRT Score

	Estimation Sample			
	CRT = 0	CRT = 1	CRT = 2	CRT = 3
	(1)	(2)	(3)	(4)
Constant (Baseline)	6.916	5.061	3.039	4.654
	(0.997)	(0.965)	(1.165)	(2.410)
Depletion Effect	-3.580**	-1.457	0.796	1.296
	(1.405)	(1.635)	(1.345)	(3.015)
Placebo Effect	-2.468*	0.251	0.096	-0.726
	(1.240)	(1.319)	(1.653)	(2.559)
Sugar Effect	-2.409*	-0.217	1.143.	-0.299
	(1.351)	(1.260)	(1.728)	(2.599)
Clusters	42	40	40	27
Observations	1890	1800	1800	1215

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster.

To rule out mood or affect as potential drivers of our sugar or placebo effects, we use the elicited mood and beverage enjoyment data from the post-drink surveys (Drink treatments) and entry surveys (Baseline).[112] First and foremost, mood is not predictive of demand in our experiment. Second, we use a specification identical to our hours-since-last-meal analysis, but replace that variable with the self-reported mood variable, and exclude individuals from the Depletion treatment (since their mood elicitation took place prior to the Stroop task). Results are in appendix Tables A3.2 and A3.3. We again find no substantive evidence that mood is related to demand for lower-score participants.[113]

To add credence to our use of the Baccalauréat exam score as a measure of cogni-

[112] Both mood and beverage enjoyment are elicited as numbers from 1 (negative) to 10 (positive).

[113] The same is true of elicited beverage enjoyment. Attempts to replicate the Ifcher and Zarghamee (2011) result by using our treatment variables as instruments for mood fail due to a lack of relevance: our treatments do not appear to affect mood.

tive ability, we present treatment effect estimates split by CRT performance instead of by Baccalauréat score.[114] As noted, these estimates should be interpreted with caution since the treatments may have affected the subjects' CRT performance, just as they affected the subjects' performance in the time-preference task. That said, consistent with our results using the Baccalauréat, we find significant effects of the treatments on time preferences only for those who failed to answer a single CRT question correctly (slightly more than 70% of these individuals are in the lower-score group). Results are presented in Table 3.8.[115]

3.4 Discussion

This paper studies the effects of prior impulse-controlling activity and sugar consumption on time preferences. A key innovation of our approach is an explicit model of intertemporal choice, which allows us to distinguish three aspects of 'patience' that might be affected by the cognitive and physiological environment: discount rates, present bias, and price sensitivity. We find that intertemporal choices are sensitive to transient features of the choice environment, but not necessarily in ways that are consistent with a willpower-based model. For example, exposure to the Stroop (1935) task prior to the elicitation of time preferences makes lower-test-score participants *more* responsive to high prices for early income. It is as if the Stroop test primed the subjects to think more carefully about their subsequent economic decisions. One interpretation is that Stroop-exposed subjects

[114] As mentioned earlier, CRT and Baccalauréat performance are positively and significantly correlated.

[115] Table 3.8's presentation of results for each possible CRT score raises the question of how our main results would change using a finer breakdown of Baccalaurat scores than whether subjects are above or below the median. To that end, appendix Figure A3.4 shows demand for early payments as cubic functions of Baccalauréat score. The results are mostly similar, although they do show a strong sugar effect at the very bottom of the cognitive ability distribution.

paid greater attention to the arbitrage opportunities available to them in capital markets, resulting in more price-sensitive choices.[116] While this result contrasts with previous experiments showing that prior impulse control reduces subjects' performance on subsequent impulse-control tasks, we note that the time-preference task in our experiment differs in important ways from the outcomes studied in those experiments. In particular, because we offer subjects a menu of choices where they choose how much to save or borrow –rather than just saying 'yes' or 'no' to a given amount of income today–, our time-preference task may be less conducive to 'snap' or thoughtless decision-making. While it is unclear which type of task is more representative of the types of consumer financial decisions – such as payday loans– we are trying to model, our surprising results still cast doubt on energy-budget-based models for a large class of financial decisions consumers make.

We also find that drinking either a sugar-free or a sugared beverage ten minutes prior to the time preference task increases patience, and does so by raising subjects' sensitivity to high prices. The magnitude of the sugar effect is significantly greater than the magnitude of the placebo effect, consistent with a role for blood glucose. However, the finding that the placebo beverage has an effect relative to the baseline, which in many specifications is larger than the marginal effect of sugar, raises questions about the importance of blood glucose relative to other situational factors.[117] It is also possible that the distinct nature

[116]Carvalho *et al.* (2013) study the effects of randomly providing individuals with a savings account on time preferences. Echoing our results, their treatment effects are concentrated on utility curvature: subjects with new accounts exhibit more linear preferences, suggesting increased sensitivity of choices to market options.

[117]Indeed, recent findings by Molden *et al.* (2012) and Sanders *et al.* (2012) show that simply rinsing one's mouth with a sugared beverage without swallowing (with no effect on blood glucose) bolsters impulse control in similar ways to ingesting sugar; related neurological evidence indicates that the sensing of the carbohydrate in the mouth activates a part of the brain that is highly sensitive to incentives (Kringelbach 2004; Chambers *et al.* 2009). Since our placebo drink did contain a very small amount of sugar –though not enough to affect blood glucose levels– this mechanism could account for our findings. An alternative explanation is

of our time-preference task accounts for the relatively small effect of sugar consumption, relative to other situational factors. In sum, our results raise questions about the importance of willpower-based models as the most appropriate way to conceptualize a large class of consumer decisions.

This said, we emphasize that our results do not imply that consumers' financial decisions are immune from situational factors; indeed our estimated treatment effects are large in magnitude. For example, the estimated changes in utility-function curvature (α) associated with our Depletion, Placebo and Sugar treatments reduce the predicted demand for a typical payday loan (two weeks at an APR of 390%, against a €1000 paycheck) from a baseline of €310 to €220, €140 and €60 respectively. The associated loan interest charges fall from €60 to €40, €20 and €10. Payday loans such as the above are considered by many to be 'predatory' in that their short-term nature takes advantage of scope insensitivity in interest rates to charge above-market rates. In these situations, our finding that all of the treatment effects operate through the intertemporal elasticity of substitution indicates that unless consumers are highly attuned to their task at hand, they may ignore substantial price differences across assets or credit payments.[118] While, as noted, all the above treatment effects are largely absent among subjects with very high cognitive abilities (corresponding to the top decile of French high school graduates), it is noteworthy that our 'lower-score'

---

that the drinks could be perceived as a reward, which might have the effect of reducing subjects' desire for an immediate secondary reward. One difficulty with this hypothesis, however, is that the drinks should act primarily on subjects' present bias, not on their price-sensitivity; the drinks should also affect both lower- and higher-score subjects, which is not what we observe.

[118]For those concerned about the external validity of our experimental measures, we point to existing literature that demonstrates a strong relationship between experimentally elicited impatience and wealth and health investment (Hastings and Mitchell, 2011), present-bias and credit card debt (Meier and Sprenger, 2010) and time discounting and credit scores (Meier and Sprenger, 2012).

results pertain to a subject pool whose cognitive ability is still well above the national mean (representing about the 50th-90th percentiles of high school graduates), suggesting the possibility of even larger effects for the population as a whole.

3.5 Acknowledgements

Chapter 3, in full, is currently being prepared for submission for publication of the material. Kuhn, Michael A.; Kuhn, Peter; Villeval, Marie Claire. The dissertation author was the primary investigator and author of this material.

3.6 Appendix

Figure A3.1: Glasses Containing either the Placebo or the Sugared Beverage

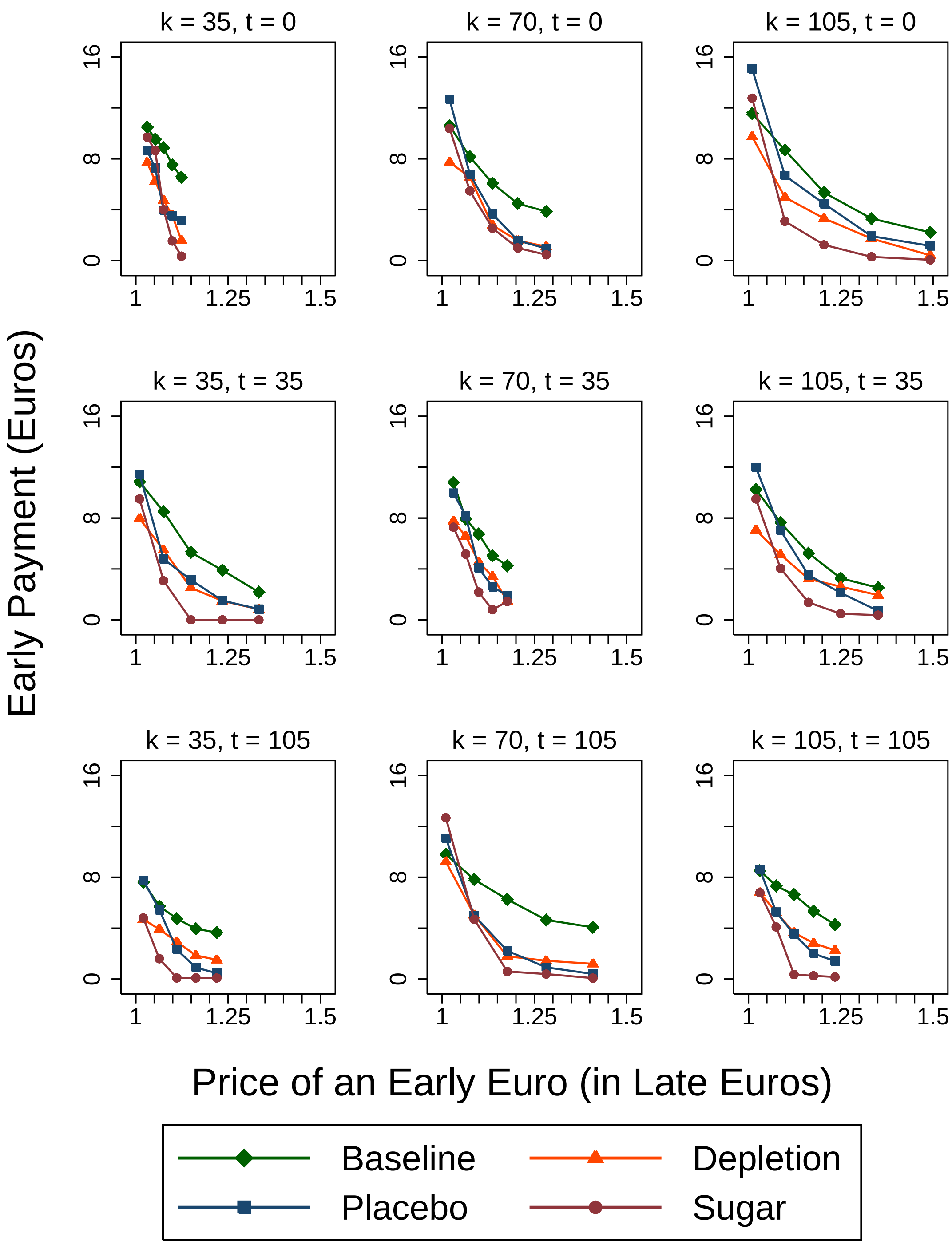


Figure A3.2: Demand Functions by Treatment, Lower-Score Sample

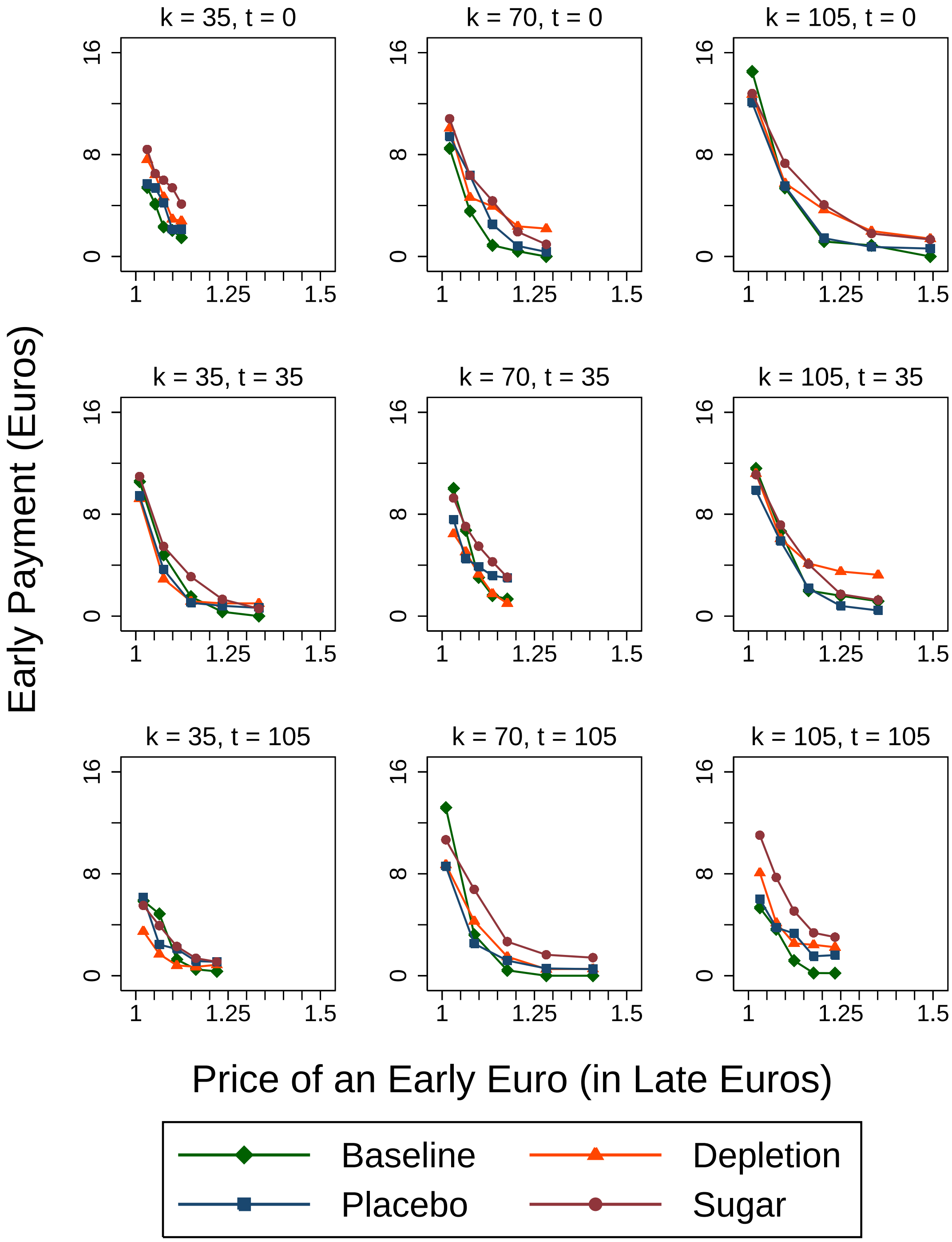


Figure A3.3: Demand Functions by Treatment, High-Score Sample

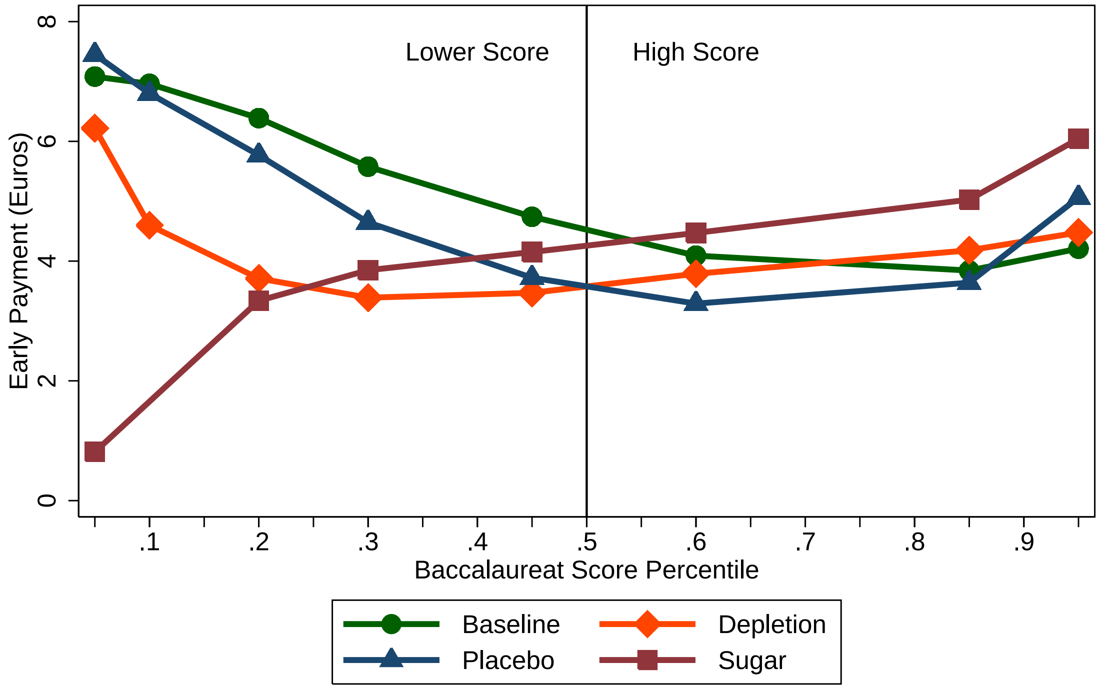


Figure A3.4: Predicted Demand using Baccalauréat Score Cubic

Predictions are from a regression of demand on a cubic in the subject's Baccalauréat Score. The data are trimmed in order to avoid estimating the polynomials on outliers.

Table A3.1: The 45 Choice Sets in the Time Preference Elicitation Task

Choice number	Early date t	Delay length k	Early value of 1 token a_t	Price of an early Euro	Annual intrest rate %	Maximum early payoff
1	0	5	0.97	1.03	36	15.52
2	0	5	0.95	1.05	65	15.20
3	0	5	0.93	1.08	100	14.88
4	0	5	0.91	1.10	141	14.56
5	0	5	0.89	1.12	189	14.24
6	5	10	0.97	1.03	17	15.52
7	5	10	0.94	1.06	36	15.04
8	5	10	0.91	1.10	59	14.56
9	5	10	0.88	1.14	85	14.08
10	5	10	0.85	1.18	116	13.60
11	15	15	0.97	1.03	11	15.52
12	15	15	0.93	1.08	28	14.88
13	15	15	0.89	1.12	47	14.24
14	15	15	0.85	1.18	70	13.60
15	15	15	0.81	1.23	96	12.96
16	0	10	0.98	1.02	11	15.68
17	0	10	0.93	1.08	44	14.88
18	0	10	0.88	1.14	85	14.08
19	0	10	0.83	1.20	139	13.28
20	0	10	0.78	1.28	208	12.48
21	5	15	0.98	1.02	7	15.68
22	5	15	0.92	1.09	32	14.72
23	5	15	0.86	1.16	64	13.76
24	5	15	0.80	1.25	103	12.80
25	5	15	0.74	1.35	154	11.84
26	15	5	0.98	1.02	23	15.68
27	15	5	0.94	1.06	82	15.04
28	15	5	0.90	1.11	164	14.40
29	15	5	0.86	1.16	278	13.76
30	15	5	0.82	1.22	432	13.12
31	0	15	0.99	1.01	4	15.84
32	0	15	0.91	1.10	37	14.56
33	0	15	0.83	1.20	83	13.28
34	0	15	0.75	1.33	144	12.00
35	0	15	0.67	1.49	231	10.72
36	5	5	0.99	1.01	11	15.84
37	5	5	0.93	1.08	100	14.88
38	5	5	0.87	1.15	246	13.92
39	5	5	0.81	1.23	479	12.96
40	5	5	0.75	1.33	845	12.00
41	15	10	0.99	1.01	5	15.84
42	15	10	0.92	1.09	51	14.72
43	15	10	0.85	1.18	116	13.60
44	15	10	0.78	1.28	208	12.48
45	15	10	0.71	1.41	339	11.36

Table A3.2: Treatment Effect on Early Payment Demand with Mood

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
Constant (Baseline, neutral mood)	5.168	7.898	2.638
	(2.506)	(3.551)	(2.112)
Placebo Effect	-1.144	-1.922	-0.412
	(0.834)	(1.163)	(0.998)
Sugar Effect	-0.878	-3.407***	1.567*
	(0.847)	(1.120)	(0.900)
Mood (-5 to 5 scale)	0.046	-0.260	0.160
	(0.438)	(0.598)	(0.424)
Mood X Placebo	-0.383	0.381	-1.008*
	(0.539)	(0.754)	(0.586)
Mood X Sugar	-0.141	0.648	-0.500
	(0.527)	(0.679)	(0.558)
Clusters	109	55	54
Observations	4905	2475	2430

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster. Mood is elicited on a 1-10 scale. We renormalize to -5 to 5 such that treatment effect estimates refer to neutral mood.

Table A3.3: Treatment Effect on Early Payment Demand with Drink Enjoyment

	Estimation Sample		
	All Subjects	Lower-Score	High-Score
	(1)	(2)	(3)
Constant (Baseline, neutral enjoyment)	5.408	6.491	3.422
	(0.674)	(0.896)	(0.681)
Placebo Effect	-1.258	-1.852*	0.004
	(0.809)	(1.037)	(1.021)
Sugar Effect	-1.054	-3.344***	1.385
	(0.803)	(1.107)	(0.862)
Placebo X Enjoyment (-5 to 5 scale)	-0.231	-0.162	-0.318
	(0.160)	(0.172)	(0.277)
Sugar X Enjoyment (-5 to 5 scale)	0.296*	0.072	0.373*
	(0.172)	(0.191)	(0.196)
Clusters	109	55	54
Observations	4905	2475	2430

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster. Enjoyment is elicited on a 1-10 scale. We renormalize to -5 to 5 such that treatment effect estimates refer to neutral enjoyment.

Table A3.4: Treatment Effects on Probability of Corner Solution Choice

	Marginal Effects from Multinomial Logit Model					
			Estimation Sample			
	All Subjects		Lower-Score		High-Score	
Corner Choice:	Sooner	Later	Sooner	Later	Sooner	Later
	(1)	(2)	(3)	(4)	(5)	(6)
Constant (Basline)	0.242	0.472	0.287	0.344	0.161	0.706
	(0.046)	(0.060)	(0.066)	(0.072)	(0.038)	(0.069)
Depletion Effect	-0.063	0.149*	-0.142*	0.217**	0.049	-0.030
	(0.057)	(0.078)	(0.084)	(0.108)	(0.058)	(0.089)
Placebo Effect	-0.055	0.111	-0.077	0.179*	-0.009	-0.033
	(0.054)	(0.080)	(0.075)	(0.096)	(0.059)	(0.111)
Sugar Effect	-0.061	0.060	-1.156**	0.307***	0.040	-0.217**
	(0.054)	(0.078)	(0.071)	(0.105)	(0.053)	(0.091)
Clusters	149		74		75	
Observations	6705		3330		3375	

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Standard Errors in parentheses, clustered by individual. 45 observations (budgets) per cluster. The multinomial logit specification estimates the effect of our treatments on the probability of choosing either the sooner or later corner solution, with respect to an interior choice (all pooled). This table presents the marginal effects of changing the treatment indicators from 0 to 1, holding the other indicators constant at 0.

3.6.1 Experimental Stimuli

We provide the following stimuli (in this order): introduction, CTB instructions, CTB decision screen, Stroop decision screen.

DrTD

Instructions

You are about to participate in an experimental session on decision-making.

The session consists of several parts. You will receive the instructions for each part after the previous part has been completed.

Part 1

Your computer screen will display a number of questions. We thank you for answering these questions with care.

Once all participants will have answered these questions, we will distribute glasses of a beverage that we will invite you to drink. Please do not drink the beverage before being expressly invited to do it.

Next, you will have to answer a few questions.

After you have answered these questions, you will have to wait for the next part. During this rest period, you are allowed to read books, newspapers or magazines. During this part and throughout the session, it is not allowed to talk to the other participants.

DrTD

Part 2

Your decisions

In this part, you will be asked to make a series of choices between payments you can receive at different dates. On each of nine decision screens, you will decide how to divide your payment for the experiment between two dates: an 'early' date and a 'late' date.

Altogether, you will make a total of 45 choices on the nine decision screens. These decision screens will be displayed in a random order. You will have the following options for payment dates:

Decide between payment <u>today</u> and payment in <u>5 weeks</u>

Decide between payment in <u>5 weeks</u> and payment in <u>15 weeks</u>

Decide between payment in <u>15 weeks</u> and payment in <u>30 weeks</u>

Decide between payment <u>today</u> and payment in 10<u> weeks</u>

Decide between payment in <u>5 weeks</u> and payment in <u>20 weeks</u>

Decide between payment in <u>15 weeks</u> and payment in <u>20 weeks</u>

Decide between payment <u>today</u> and payment in <u>15 weeks</u>

Decide between payment in <u>5 weeks</u> and payment in <u>10 weeks</u>

Decide between payment in <u>15 weeks</u> and payment <u>25 weeks</u>

On each decision screen, we will provide you with the exact calendar dates of the above payments, so you know exactly which decision you are making. Today's date appears in green, the early payment date appears in blue and the late payment date appears in red.

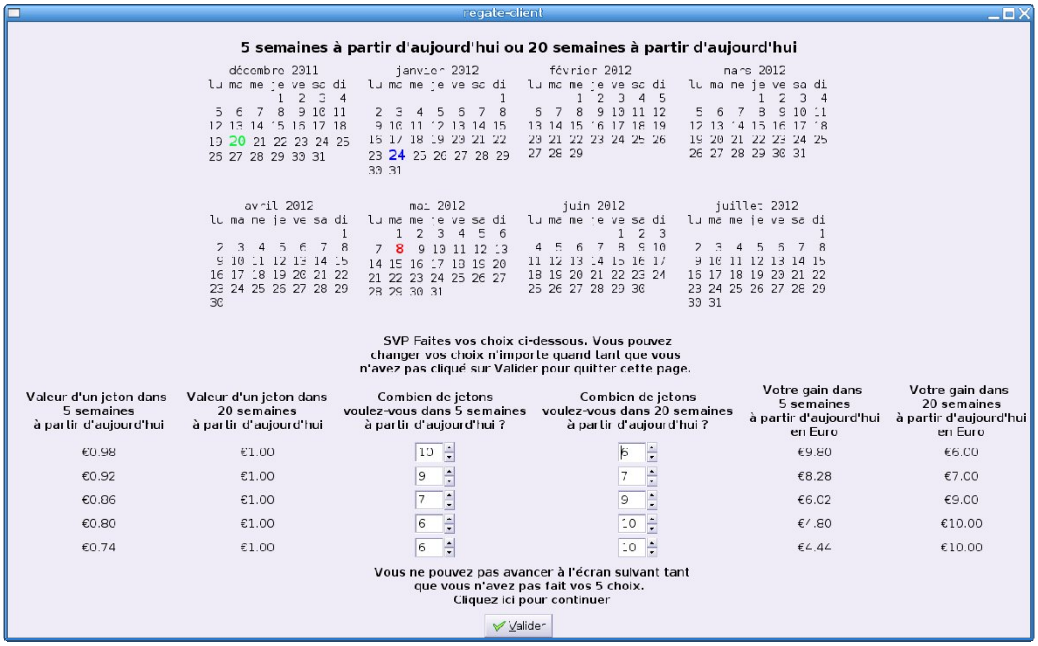
You will be given 16 tokens to divide in each choice, but *the value of a token changes from choice to choice.* The real money payments associated with your token choices will be automatically calculated for you to see as you make your decisions.

To make your decisions, you can enter a number for the early payment (or the late payment) and move the up and down arrows. The box corresponding to the late payment (or the early payment, respectively) will be automatically updated by a number indicating the difference between 16 and the tokens assigned to the other date of payment.

Once you have completed a set of five decisions, you must press the "Validate" button to move to the next decision screen.

Below is an example of a decision screen.

DrTD



Your payment

At the end of the session, the computer program will randomly select one of the 45 decisions you made to be your earnings from participating in this experiment.

In addition, you will receive a €5 participation payment that will be split up into two payments of €2.50: one to go along with your earnings at the early and late dates associated with the randomly selected decision.

This means that you will not be paid in cash today. You will be paid by checks that will be mailed to you at the address you will indicate on the envelopes on your desk. We will mail the envelopes at the dates corresponding to the randomly selected decision.

For example, if the selected decision indicates that you have chosen x tokens today and y tokens in 10 weeks, we will mail the first check today and the second check in 10 weeks from today.

Remember that each decision could be the one that counts! Treat each decision as if it could be the one that determines your payment.

If you have any question on these instructions, please raise your hand and we will answer your questions in private.

DrTD

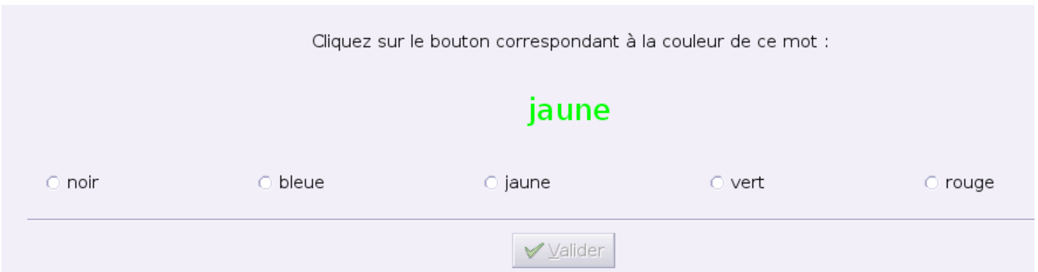
Part 3

In this part, you will be presented with a series of color words (black, blue, yellow, green, red). These words will appear in different colors, sometimes matching the word (e.g., the word blue, written in blue), and sometimes not matching the word (e.g., the word blue, written in red).

Your job is to indicate, as quickly and accurately as possible, the color in which the word is written, whether or not that matches the word itself. Click the button that matches the color of the word. Try not to pay attention to the word, but just the color.

This task will last for six minutes.

Example :



In this example, the correct answer is « green ».

3.7 References

L. Aarøe and M. Bang Petersen. Hunger games: Fluctuation in blood glucose levels influence support for social welfare. *Psychological Science*, 24(12):2250–2556, 2013.

J. Andreoni and C. Sprenger. Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–3356, 2012.

J. Andreoni, M.A. Kuhn, and C. Sprenger. On measuring time preferences. Working paper, UC San Diego, 2013.

N. Augenblick, M. Niederle, and C. Sprenger. Working over time: Dynamic inconsistency in real effort tasks. Working paper, NBER No. 18734, 2013.

R.F. Baumeister, E. Bratslavsky, M. Muraven, and D.M. Tice. Self-control depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5): 1252–1265, 1998.

D.J. Benjamin, S.A. Brown, and J.M. Shapiro. Who is "behavioral"? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 2014. Forthcoming.

B.D. Bernheim and A. Rangel. Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–1590, 2004.

D. G. Blanchflower, A.J. Oswald, and S. Stewart-Brown. Is psychological well-being linked to the consumption of fruit and vegetables? Working paper, NBER No. 18469, 2012.

B. Briers, M. Pandelare, S. Dewitte, and L. Warlop. Hungry for money: The desire for caloric resources rsources increases the desire for financial resource and vice-versa. *Psychological Science*, 17(11):939–943, 2006.

A. Bucciol, D. Houser, and M. Piovesan. Temptation and productivity. *Journal of Economic Behavior and Organization*, 78(1):126–136, 2011.

N. Burger, G. Charness, and J. Lynham. Field and online experiments on self-control. *Journal of Economic Behavior and Organization*, 77(3):393–404, 2011.

M. Callen, M. Isaqzadeh, J.D. Long, and C. Sprenger. Violence and risk preference: Experimental evidence from afghanistan. *American Economic Review*, 2013. Forthcoming.

G.D. Carroll, J.J. Choi, D. Laibson, B.C. Madrian, and A. Metrick. Optimal defaults and active choices. *Quarterly Journal of Economics*, 124(4):1639–1674, 2009.

L. Carvalho, S. Prina, and J. Sydnor. The effects of savings on risk-taking and intertemporal choice behavior: Evidence from a randomized experiment. Working paper, 2013.

C.F. Chabris, D. Laibson, and J.P. Schuldt. Intertemporal choice. In S. Durlauf and L. Blume, editors, *New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

E.S. Chambers, M.W. Bridge, and D.A. Jones. Carbohydrate sensing in the human mouth: Effects on exercise performance and brain activity. *Journal of Physiology*, 587(8):1779–1794, 2009.

S. Danziger, J. Levav, and L. Avanaim-Pesso. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17):6889–6892, 2011.

C.N. DeWall, R.F. Baumeister, T.F. Stillman, and M.T. Gailliot. Violence restrained: Effects of self-regulation and its depletion on aggression. *Experimental Social Psychology*, 43(1):62–76, 2007.

D.L. Dickinson, T. McElroy, and N. Stroh. The impact of glucose administration on bayesian vs. heuristic based choice. Working paper, Appalachian State University, 2013.

T. Dohmen, A. Falk, D. Huffman, and U. Sunde. Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260, 2010.

R.T. Donohoe and D. Benton. Blood glucose control and aggressiveness in females. *Personality and Individual Differences*, 26(5):905–911, 1999.

D. Dotter. Breakfast at the desk: The impact of universal classroom breakfast programs on academic performance. Working paper, UC San Diego, 2013.

S. Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, 2005.

D. Fudenberg and D.K. Levine. A dual-self model of impulse control. *American Economic Review*, 96(5):1449–1476, 2006.

D.C. Funder and J. Block. The role of ego-control, ego-resiliency, and iq in delay of gratification in adolescence. *Journal of Personality of Social Psychology*, 57(6):1041–1050, 1989.

M.T. Gailliot and R.F. Baumeister. The physiology of willpower: Linking blood glucose and self-control. *Personality and Social Psychology Review*, 11(4):303–327, 2007.

M.T. Gailliot, R.F. Baumeister, C.N. DeWall, J.K Maner, E.A. Plant, D.M. Tice, L.E. Brewer, and B.J. Schmeichel. Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92 (2):325–336, 2007.

M.T. Gailliot, B.M. Peruche, E.A. Plant, and R.F. Baumeister. Stereotypes and prejudice in the blood: Sucrose drinks reduce prejudice and stereotyping. *Journal of Experimental Social Psychology*, 45(1):288–290, 2009.

E.H. Gordijn, I. Hindriks, W. Koomen, A. Dijksterhuis, and A. Van Knippenberg. Consequences of stereotype suppression and internal suppression motivation: A self-regulation approach. *Personality and Social Psychology Bulletin*, 30(2):212–224, 2004.

B. Greiner. An online recruitment system for economic experiments. *Forschung und wissenschaftliches Rechnen*, 63:79–93, 2004.

T.A. Hare, J. Malmaud, and A. Rangel. Focusing attention on the health aspects of food changes value signals in vmpfc and improves dietary choice. *Journal of Neuroscience*, 31(30):11077–11087, 2011.

A. Harris, T.A. Hare, and A. Rangel. Temporally dissociable mechanisms of self-control: Early attentional filtering versus late value modulation. *Journal of Neuroscience*, 33(48): 18917–18931, 2013.

J.S. Hastings and O.S. Mitchell. How financial literacy and impatience shape retirement wealth and investment behaviors. Working paper, NBER No. 16740, 2011.

J.M. Hinson, T.J. Jameson, and P. Whitney. Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(2): 298–306, 2003.

J. Ifcher and H. Zarghamee. Happiness and time preference: The effect of positive affect in a random-assignment experiment. *American Economic Review*, 101(7):3109–3129, 2011.

D. Kahan, J. Polivy, and C.P. Herman. Conformity and dietary disinhibition: A test of the ego-strength model of self-regulation. *International Journal of Eating Disorders*, 32(2): 165–171, 2003.

C. Kirby, G. Winston, and M. Santiesteban. Impatience and grades: Delay-discount rates correlate negatively with college gpa. *Learning and Individual Differences*, 15(3):213–222, 2005.

B. Koszegi and A. Szeidl. A model of focusing in economic choice. *Quarterly Journal of Economics*, 128(1):53–104, 2013.

M.L. Kringelbach. Food for thought: Hedonic experience beyond homeostasis in the human brain. *Neuroscience*, 126(4):807–819, 2004.

D. Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):807–819, 1997.

C.M. MacLeod. Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin*, 109(2):163–203, 1991.

N.L. Mead, R.F. Baumeister, F. Gino, M.E. Schweitzer, and D. Ariely. Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45(3):594–597, 2009.

S. Meier and C. Sprenger. Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210, 2010.

S. Meier and C. Sprenger. Time discounting predicts creditworthiness. *Psychological Science*, 23(1):56–58, 2012.

W. Mischel, E.B. Ebbesen, and A. Raskoff Zeiss. Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, 21(2):204–218, 1972.

D.C. Molden, C.M. Hui, A.A. Scholer, B.P. Meier, E.E. Noreen, P.R. D'Agostino, and V. Martin. Motivational versus metabolic effects of carbohydrates on self-control. *Psychological Science*, 23(10):1137–1144, 2012.

M.R. Muraven and R.F. Baumeister. Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2):247–259, 2000.

M.R. Muraven, R.L. Collins, S. Shiffman, and J.A. Paty. Daily fluctuations in self-control demands and alcohol intake. *Psychology of Addictive Behaviors*, 19(2):140–147, 2005.

T. O'Donoghue and M. Rabin. Doing it now or later. *American Economic Review*, 89(1): 103–124, 1999.

A.G. Oswald, E. Proto, and D. Sgroi. Happiness and productivity. *Journal of Labor Economics*, 2014. Forthcoming.

E. Ozdenoren, S.W. Salant, and D. Silverman. Willpower and the optimal control of visceral urges. *Journal of the European Economic Association*, 10(2):342–368, 2012.

J.A. Richeson and J.N. Shelton. When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, 14(3):287–290, 2003.

J.A. Richeson and S. Trawalter. Why do interracial interactions impair executive function? a resource depletion account. *Journal of Personality and Social Psychology*, 88(6):934–947, 2005.

J.A. Richeson, S. Trawalter, and J.N. Shelton. African americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition*, 23(4):336–352, 2005.

A. Rustichini, C. DeYoung, J. Anderson, and S.V. Burks. Toward the integration of personality theory and decision theory in the explanation of economic and health behavior. IZA Discussion Paper 6750, 2012.

M.A. Sanders, S.D. Shirk, C.J. Burgin, and L.L. Martin. The gargle effect: Rinsing the mouth with glucose enhances self control. *Psychological Science*, 23(12):1470–1472, 2012.

N.A. Shamosh and J.R. Gray. Delay discounting and intelligence: A meta-analysis. *Intelligence*, 36(4):289–305, 2008.

H.M. Shefrin and R.H. Thaler. The behavioral life-cycle hypothesis. *Economic Inquiry*, 26 (4):609–643, 1988.

Y. Shoda, W. Mischel, and P.K. Peake. Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, 26(6), 1990.

J.R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662, 1935.

T.S. Stucke and R.F. Baumeister. Ego depletion and aggressive behavior: Is the inhibition of aggression a limited resource? *European Journal of Social Psychology*, 36(1):1–13, 2006.

K.D. Vohs and T.F. Heatherton. Self-regulatory failure: A resource-depletion approach. *Psychological Science*, 11(3):249–254, 2000.

X.T. Wang and R.D. Dvorak. Sweet future: Fluctuating blood glucose levels affect future discounting. *Psychological Science*, 21(2):183–188, 2010.