Which Measures of Time Preference Best Predict Outcomes?
Evidence from a Large-Scale Field Experiment

Stephen V. Burks
Jeffrey Carpenter
Lorenz Götte
Aldo Rustichini

DISCUSSION PAPER SERIES

Which Measures of Time Preference Best Predict Outcomes? Evidence from a Large-Scale Field Experiment

Stephen V. Burks
University of Minnesota Morris and IZA

Jeffrey Carpenter
Middlebury College and IZA

Lorenz Götte
University of Lausanne and IZA

Aldo Rustichini
University of Minnesota

ABSTRACT

Which Measures of Time Preference Best Predict Outcomes? Evidence from a Large-Scale Field Experiment[*]

Economists and psychologists have devised numerous instruments to measure time preferences and have generated a rich literature examining the extent to which time preferences predict important outcomes; however, we still do not know which measures work best. With the help of a large sample of non-student participants (truck driver trainees) and administrative data on outcomes, we gather four different time preference measures and test the extent to which they predict both on their own and when they are all forced to compete head-to-head. Our results suggest that the now familiar (β, δ) formulation of present bias and exponential discounting predicts best, especially when both parameters are used.

NON-TECHNICAL SUMMARY

A great deal of evidence suggests that when waiting increases the size of the monetary reward, many economic decisions and outcomes are affected by the degree to which individuals are willing (or not) to wait for future payments. Using a large sample of non-student subjects (trainee truckers) who took part in an extensive set of economic field experiments, we compare the effectiveness of four ways of measuring such "time preferences" at predicting real life outcomes for these subjects. The measurements are two standard parameters ("beta" and "delta") derived from choices over future payments, surveyed impatience, and performance on a waiting task. The outcomes are smoking, credit score, body mass index, and on-the-job success. Though each has strengths and weaknesses, overall the beta-delta model does best.

Corresponding author:

Stephen V. Burks
Division of Social Science
University of Minnesota Morris
600 East 4th Street
Morris, Minnesota 56267
USA
E-mail: svburks@morris.umn.edu

1 Introduction

Since the early work of Irving Fisher (1930) and Paul Samuelson (1937), economists have placed a lot of theoretical weight on time preferences to explain individual choices across a variety of domains from finance to bargaining or even lifestyle choices that affect one's health. At around the same time psychologists also became interested in time preferences but tended to focus on the ability of people to delay gratification, control their impulses and the links between these abilities and personality. The culmination of all this interest is a rich variety of ways to measure time preferences. Within economics it is now standard, via an incentivized experiment, to elicit a simple discount factor, $0 \leq \delta \leq 1$, for future payoffs, although if present bias is thought to be important, the difference in how discounting occurs when today is involved is also typically captured using $0 \leq \beta \leq 1$ (Laibson, 1997; Harrison et al., 2002). Together, these parameters now formulate the (β, δ) model of preferences (Frederick et al., 2002). In psychology, the delay of gratification literature was heavily influenced by the impulsivity work of Walter Mischel, who showed that a simple experiment in which young children were asked to choose between one cookie (or marshmallow) now and two in fifteen minutes could predict achievement later in life (Mischel et al., 1989). In addition, along the way in both disciplines surveys have been developed to capture not only the more common elements of time preference like δ (e.g., the Health and Retirement Survey; Barsky et al., 1997), β (Ameriks et al., 2007) and impulsivity (Dickman, 1990), but some surveys have also embedded time preferences in broader theories of personality (e.g., the Jenkins Activity Survey for Type A personality; Jenkins et al., 1967).

Given all these ways to measure time preference, we are interested, as others have been in the past, in the important question of which ways best predict outcomes. For example, there is a rich tradition in economics of measuring associations between discount factors and outcomes like human capital accumulation (Eckel et al., 2005) or savings (Ashraf et al., 2006). Using a different measure but asking a similar question, psychologists have also been interested in how impulsiveness predicts school achievement (e.g., Shoda et al., 1990). Given the current work in both disciplines on the prevealence of hyperbolic discounting, many recent papers have studied the ability of both δ and β to predict outcomes. As just three examples, Meier and Sprenger (2010) find a much stronger relationship between present bias and credit card debt than between this sort of debt and a traditional measure of one's discount factor, Mitchell (1999) shows that smokers appear less patient on both a hypothetical discounting task and a surveyed impulsivity scale, and Zhang and Rashad (2008) report that surveyed "willpower" correlates with body mass index (BMI), especially for men. While there is considerably more research in this vein (e.g., Picone et al., 2004 or

1

Eckel et al., 2007), one thing about the existing literature that is particularly conspicuous is that in most cases time preferences are measured using only one method and, as a result, no proper "horse race" has been conducted to see if one method robustly predicts better than the others.

We report the results of an experiment in which we gather four measures of time preference designed to be representative of the current methods (i.e., δ, β, impulsivity and surveyed impatience) and test the extent to which these measures predict important health outcomes like smoking and BMI, credit scores, and subsequent job related outcomes. Aside from having all four measures of time preference and being able to let them compete "head-to-head" for the first time, there are a number of other factors that make our study unique. Instead of relying on university students who rarely face all the important decisions that we study, our participants were older and had more life experience. Our sample is also much larger than the typical lab experiment. We were able to gather preference and outcome data, along with a large number of other regressors, from more than one thousand participants. While previous studies have often relied on surveyed outcomes, all but one of our outcome variables come from adminstrative data and therefore are not subject to any self-reporting biases. Lastly, it is common in this literature for causation to be ambiguous, largely because time preferences and outcomes are collected coterminously and therefore the direction of causation is unclear. While this is true for three measures that we use to replicate previous work (smoking, BMI and credit scores), we were also able to gather subsequent job performance measures that, together with our rich set of controls, allow us to be more confident that the relationships that we estimate are causal.

Our results suggest that all of the four measures of time preference have some predictive power. For example, run in isolation our behavioral measure of impulsivity predicts smoking, β predicts credit scores, and δ predicts not just leaving the job but going absent without leave (AWOL) to list just a few of the interesting correlations. However, when they are all allowed to compete for variation in the same regression model with a number of controls, it appears that the quasi-hyperbolic (β, δ) formulation predicts best. Specifically, we find that neither surveyed impatience nor impulsivity predict outcomes very well. At the same time, however, we find that present biased participants (i.e., those with low βs) are more likely to smoke, leave the job for any reason, and wash out of job training. In addition, participants with low discount rates (i.e., high δs) are also less likely to smoke, have better credit and are less likely to go AWOL from the job.

We proceed by first describing our measures of time preference and our participants in Section 2. In Section 3 we examine our measures of time preference in detail by looking for demographic correlates and by assessing the extent to which the instruments measure a

2

common trait. Our main results are presented in Section 4 where we begin by looking at the simple correlations between the measures of time preference and six outcomes. We then add controls and force all four measures to compete for variation in the outcome measures. At the end of this section we consider a few robustness checks. Section 5 concludes the paper with a summary.

2   Eliciting Time Preferences

Between December 2005 and August 2006 we gathered data from 1,069 truck driver trainees at one of the training facilities of a large midwestern motor carrier firm. Ninety-one percent of the firm's trainees to whom we offered the opportunity participated. There were between 18 and 30 participants in each group; each group took part in two two-hour long sessions. In addition to a show-up fee of $20 ($10 paid at the beginning of each session), all parts of the experiment (except the questionnaires) were incentivized. In the end, participants earned between $21 and $168 with an average of $53. A more complete account of the design and the context of the broader project can be found in Burks et al. (2008). We continue by providing the details of the time preference measures that we gathered and by summarizing the other survey and administrative data that we use in the present paper.

2.1   Measures of Quasi-hyperbolic Discounting

Our instrument to measure δ and β was constructed similarly to many of those surveyed in Frederick et al., (2002). Participants made a total of 28 binary choices between a sooner smaller amount of money and a later but larger amount. The amount of the later larger payment was fixed for each choice at $80 and the amount of the sooner smaller payment stepped down from $75 to $45 in $5 decrements. In principle, participants would begin by choosing the early payment until the total decrement is large enough and then they flip over and choose the later payment.

To identfy present-bias separately from discount factors, the choices were split into four blocks of seven choices each and organized as follows: (i) today versus tomorrow, (ii) today versus five days from today, (iii) two days from today versus nine days from today and (iv) two days from today versus thirty days from today. Because we were explicitly interested in measuring present-bias, we decided to not employ a front-end delay (a la Coller and Williams, 1999) longer than the duration of the experiment for the first two blocks. Considering we were sanctioned by the firm's management to conduct our research, the typical rationale for adding such a delay - that not believing you will be paid in the future makes one seem more

impatient when today is involved - was much less likely to be a problem; all participants could be sure that they would be paid.

To pay participants for this part of the experiment, 28 numbered chips were placed in a bowl and one was drawn at random to determine the choice that would count. Two more chips were then drawn from a different bowl that included one chip for each participant in the session to identify two people at random to pay. Payments were either made in person or via mailed official bank checks.

We are able to use the different delays to identify individual-specific estimates of δ_i and whether or not today was involved to identify estimates of β_i. Specifically, we adopt the model of (β, δ)-preferences proposed by Phelps and Pollack (1968) and Laibson (1997). We measure the amount x_i such that the individual is indifferent between receiving it now and receiving \$80 t periods (measured in days) from now. Indifference implies

$$u(x_i) = \beta \delta^t u(80). \tag{1}$$

If the choice is between receiving an amount x_i s periods from now or \$80 $t + s$ periods from now, indifference then implies

$$u(x_i) = \delta^t u(80). \tag{2}$$

Taking logs of equations (1) and (2), we get

$$\log u(x_i) - \log u(80) = \log \beta + t \log \delta \tag{3}$$

if the choice involves today (as in our choice blocks 1 and 2), and

$$\log u(x_i) - \log u(80) = t \log \delta \tag{4}$$

if it doesn't (as in our choice blocks 3 and 4). We assume that u is approximately linear over the relevant range and then use ordinary least squares and the following specification to estimate β_i and δ_i for individual i in choice block $k = 1, 2$

$$\log x_{i,k} - \log 80 = \log \beta_i + t_k \log \delta_i + e_{i,k}. \tag{5}$$

For choice block $k = 3, 4$ we use

$$\log x_{i,k} - \log 80 = t_k \log \delta_i + e_{i,k}. \tag{6}$$

In each case $x_{i,k}$ is the amount at which the individual switched to the future payment in

4

block k, t_k is the delay (in days) of the larger payment, and $e_{i,k}$ is a mean-zero error term.

2.2   A New Measure of Impulsivity

To try to capture the essence of the Mischel delay of gratification experiments (Mischel et al., 1989) we decided to design a new, adult version of the task in which participants were forced to trade off unproductive waiting time against monetary compensation (instead of treats).

As described above, the entire experimental session was split into two segments, each about two hours long, between which there was a break of between 10 and 20 minutes. The impulsivity experiment came at the end of the first segment. In this experiment, which we call the (big) red button, participants were forced to sit at their stations and do nothing for 10 minutes while a timer on the computer screen counted down from 600 (in seconds). They were not allowed to talk, surf the internet, read or do anything other than wait quietly. The instructions on the screen read, "If you stay another 10 minutes, we will pay you \$5 on top of what you have already earned. You do not have to stay, however. On the next screen, there is a red button that you can click. Each click of the button shortens the amount of time you have to wait until this activity is over."

The instructions then stated the "cost of clicking" function: the first click of the red button lowered one's payoff by \$1 but reduced the waiting time by 5 minutes. The second click of the button cost an additional dollar but reduced the wait time by another 3 minutes and the last click of the button again cost a dollar but reduced the wait 2 more minutes to zero. We take the number of clicks as a measure of impulsivity.

It is important to note that the activity directly preceding the red button experiment was a demographic survey that participants completed at their own pace. This ordering of the activities was done purposely so that people would start the red button experiment at different times and any peer effects would be minimized. When seeing someone leave, the other participants could not tell whether that person had taken a relatively long time on the demographics and then clicked the red button or was done relatively early with the demographics and then waited the full 10 minutes in the red button experiment.

2.3   Surveyed Impatience

Based on the questions asked in a variety of surveys in the psychology literature meant to capture short term impatience, we added six questions from the World Health Organization's self-reported scale for adults to the survey segment of our experiment. Participants were asked to offer responses on a 5-point Likert scale (never, rarely, sometimes, often, very often)

and the exact wording of the questions was as follows: (i) How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done? (ii) How often do you have difficulty getting things in order when you have to do a task that requires organization? (iii) How often do you have problems remembering appointments or obligations? (iv) When you have a task that requires a lot of thought, how often do you avoid or delay getting started? (v) How often do you fidget or squirm with your hands or feet when you have to sit down for a long time? (vi) How often do you feel overly active and compelled to do things, like driven by a motor?

To summarize the responses to these questions we dichotomize the responses, add up the number of times a participant expressed extreme views (i.e., responding very often) and use factor analysis on the dichotomized responses. The resulting six-question impatience scale seems to have some consistency (Chronback's α equals 0.54 which is not too small considering the number of questions) and there appears to be a common factor because the analysis results in an eigenvalue of 1.83 which exceeds the standard threshold of 1.

2.4   Our Participants

The top portion of Table 1 provides descriptive statistics on our participants. In addition to a demographic survey, each participant completed two cognitive ability assessments. The first was a computerized version of the non-verbal IQ instrument discussed in (Raven et al., 2003) in which participants were asked to complete patterns. The second, our numeracy instrument, was developed by the Educational Testing Service and asked people to add, subtract and compare numbers. It was one half of the adult quantitative literacy exam administered in the format provided by the ETS. The average participant scored 45 out of 60 in IQ, 8 out of 12 in numeracy. In addition, there is considerable variation in the cognitive ability of our participants: approximately one quarter of the participants did not correctly complete more than two-thirds of the IQ patterns or answer more than half of the numeracy questions correctly.

Considering the more standard demographics, our average participant was middle-aged (37 years old), more than half (57%) had some schooling beyond high school, about half (48%) were married and very few minorities enter training at the upper Midwest location where we collected data: just 14% were African American and only 3% said that English was a second language. We also have data on the previous work and earning experiences of our participants. On average, participants had been unemployed for slightly more than two months (2.23) in the previous two years, 70% reported having family incomes less than thirty thousand dollars per year and almost half (45%) reported monthly expenditures greater than

$1500. Lastly, more than a quarter of our participants (28%) had spent some time in the military.

3   Time Preferences, their Determinants and the Inter-correlations

The second section of Table 1 summarizes the results from our time preference instruments. Because factor analysis suggested that we should focus our attention on the extreme responses from our six-question impatience scale, the mean of adding the indicators for each question is rather low (0.21). At the same time, 15% of respondents had one or more extreme views so there is some variation in the scale.

Also in Table 1 we see that our estimates of present-bias vary from being extremely impatient when today is concerned ($\beta_i = 0.56$) to actually being slightly future-, not present-, biased ($\beta_i = 1.07$). Despite the variation in β_i, the modal estimate is for people to show no bias and be indistinguishable from exponential discounters (i.e., $\beta_i = 1$). The range of δ_i is much more confined because we calculate daily discount factors. That said, there is still considerable variation and two prominent modes arise, one near $\delta_i = 0.976$ which is consistent with rarely choosing to wait and another at $\delta_i = 1$ indicative of people who always waited.

In Table 2 we examine four discounting types. In the first cell (i.e., where $\delta = \beta = 1$) we place 52% of our participants because our estimates indicated that they were statistically indistinguishable from people who were perfectly patient and not present-biased (i.e., for these people we could not reject the null that $\delta_i = 1$ and $\beta_i = 1$ at the 10% level). Here the mean daily discount factor is 0.989 and the mean estimated level of present bias is 0.921. Along the diagonal in Table 2 we find that 6% of our participants had estimated βs and δs that were both significantly less than 1. For these people the mean δ is 0.986 and the mean β is 0.826. We also find that 9% of our participants demonstrate significant present bias but are otherwise patient (i.e., $\delta = 1$ but $\beta < 1$) and that the remaining 33% are not present biased ($\beta = 1$) but appear to discount outcomes significantly ($\delta < 1$).

One quarter of our participants clicked the red button at least once. The distribution of clicks is such that 15% of people clicked only once, 3% clicked twice and almost 6% clicked three times. In terms of waiting time, Table 1 indicates that, indeed, most people waited the entire 10 minutes because the average time waited is 525 seconds. At the other end of the spectrum, the average time waited by the people who clicked three times was only 75 seconds and 17% of these people waited less than 30 seconds, in total.

In Table 3 we use ordinary least squares, probit and negative binomial regressions (with robust standard errors) to estimate the impact of the participant demographics on the elicited measures of time preference. For convenience, we report standardized marginal effects for the continuous variables. In the first column, we see that only schooling above the high schoole level and military service predicts surveyed impatience (here the dependent variable is the factor score). Those participants with some secondary education respond less patiently in the survey ($p < 0.10$) and those who had previously served in the military were more patient ($p < 0.01$).

The second column of Table 3 indicates that older participants and those with higher cognitive skills are less present-biased (i.e., have βs closer to 1).[1] However, African Americans and married participants appear significantly more present-biased. A standard deviation increase in numeracy is associated with a 0.185 standard deviation increase in β ($p < 0.01$). The effects of age and education are more modest. A standard deviation increase in age is associated with a 0.091 standard deviation increase in β ($p < 0.01$). Similar associations arise in column 3 for the discount factor. Here a standard deviation increase in numeracy is associated with an increase in the daily discount factor of 0.105 standard deviations ($p < 0.01$). Compared to the β estimates, the effect of age is even stronger. A standard deviation increase in age correlates with a 0.163 standard deviation increase in δ ($p < 0.01$). The effect of education is also stronger on δ ($p < 0.01$). Although the effects are significant, the magnitudes are small on a few other demographics. Considering English as a second langauge and having relatively high monthly expenditures are both associated with being more patient and being African American is associated with being less patient ($p < 0.01$).

In the last two columns of Table 3, we see that cognitive skills, age and education are factors that also influence the number of red button clicks. People with higher cognitive skills are significantly less likely to click at all ($p < 0.01$) and click less often ($p < 0.01$). However, unlike β and δ, older people are more, not less, likely to click ($p < 0.01$) and the effect is substantial. Similar to our numeracy results, we also find that having more education is significantly associated with clicking less ($p = 0.01$).

As one can readily see in Table 3, there are a few factors that seem to commonly affect our measures of time preference (e.g., cognitive ability, education and age) and, as a result, it is also likely that our measures are related and measure slightly different versions of a common trait. However, the correlations do not always follow the same pattern. While higher cognitive ability and more education are consistently associated with more patience as measured by β, δ and the red button, when we consider age the effects bifurcate. Only with β and δ is age associated with more patience. To examine the intercorrelations of our measures

---

[1] A similar effect of age on being present-biased has been reported in Ameriks et al. (2007).

more closely, we calculated the raw rank correlations between β, δ, the factor score from the survey measure of impatience and whether or not one clicked the red button. Importantly, although the resulting correlations are low, they are all consistent with a common latent trait: not clicking is positively correlated with both β and δ ($\rho = 0.066$, $p = 0.01$ and $\rho = 0.071$, $p = 0.01$, respectively) and it is negatively correlated with the survey factor score ($\rho = 0.042$, $p = 0.16$).

As further evidence that clicking the red button also measures some aspect of time preference, in Table 4 we report the results of regressing the number of red button clicks on the other three measures of time preference, both one at a time in columns (1)-(3) and altogether in column (4). Individually, each measure is significantly associated with red button clicks. As one becomes less present biased (i.e., as $\beta \to 1$) or more patient ($\delta \to 1$) people click the button less ($p < 0.01$ and $p = 0.02$, respectively) and those people who appear less patient on the survey also click more ($p = 0.10$). However, what might be the most interesting result in Table 4 is that when all three other measures of time preference are included in the same regression, only present bias remains significant. To us this indicates what we hoped for - red button clicks are likely to measure the impulsivity inherent in other measures of present bias.

4   Time Preferences and Outcomes

We split our analysis of the predictive ability of the different time preference measures into two parts. The first part examines outcome variables that are more common in the literature and were collected at the same time as our experiments by the firm that we worked with. Although we have a rich set of controls and these outcomes are not likely to be biased by self-reporting (with the possible exception of our height and weight measures) because the firm gathered the data[2], it is hard to be confident that, like the previous literature, we are capturing more than associations. In the second part of the analysis, however, we can be much more confident that our we are estimating causal relationships because our outcomes were gathered in the months following the experiments.

4.1   Coterminous Outcomes

Returning to Table 1, in the third section we summarize the data that we collected on coterminous outcomes. While at the training program, the firm put up trainees at a local

---

[2]Karlan and Zinman (2008), for example, illustrate the importance of not relying exclusively on self-reports of behavior.

motel. We knew whether or not a participant smoked from his or her motel reservation preference - 46% did at the time of our study. Height and weight data were collected as categorical variables in our demographic survey from which we could calculate body mass index by assigning values and using the standard formula: $(703 \times weight)/(height^2)$. By this measure, 42% of our participants had scores above 30 and could, therefore, be considered obese. Because the trainees effectively post a performance bond at the beginning of the course by signing a training contract which states that they will pay back the market value of training (between \$3500 and \$5000 at the time of data collection) if they do not stay on the job for a year, the employer was also able to collect the credit scores of the trainees at our request. Many of our participants did not have "good" credit. The average FICO score in our sample is 588 which is below the standard sub prime cutoff of 620 (Keys et al., 2010).

We utilize the administrative data to test if our measures of time preference can predict the incidence of smoking, credit scores and BMI. As a first pass, consider Figure 1 in which surveyed impatience, β, δ, and the number of red button clicks (aka impulsivity) are split at their medians to demonstrate the "raw" effects of each measure on normalized outcomes. According to the first graph from the left, surveyed impatience, by itself, does not have a lot of predictive power. Those scoring relatively high on impatience are no more or less likely to smoke ($p = 0.61$), have about the same credit worthiness ($p = 0.99$), and BMI ($p = 0.39$) as those who score low. The other three measures of time preference predict better, however. In the second graph we see that present biased participants (those with lower than median βs) are significantly more likely to smoke ($p < 0.01$), and have significantly worse credit scores ($p < 0.01$). The exact same story holds for the discount factor in the third graph. Again, the less patient (i.e., those with lower δs) are more likely to smoke and have worse credit ($p < 0.01$ in both cases). Contrary to Zhang and Rashad (2008), we find that none of our measures predict BMI.

As one can see from Figure 1, there are a number of significant correlations between the coterminous outcomes and our four time preference measures; however, we are most interested in running a proper race to see which measures survive when all are forced to compete for the variation in outcomes. In Table 5 we present estimates of the associations between all the time preference measures and smoking, credit scores and BMI, including a full set of controls.

Column (1) lists the marginal effects after a probit estimation for smoking. We find significant effects of both discounting and present bias: standard deviation increases in both β and δ are associated with a reduction of approximately a tenth of a standard deviation in the likelihood of smoking ($p < 0.05$ in both cases).[3] We find, however, no significant

---

[3]Chabris et al., (2008) find a similar result - a significant link between discounting and self-reports of

relationship between smoking and our impatience survey or the number of red button clicks. As for demographics, we find that African American trainees are 24.3% less likely to smoke ($p < 0.01$), married drivers are 6.6% less likely to smoke ($p < 0.10$), that those trainees with high monthly expenditures are also 8.2% less likely to smoke ($p < 0.10$), and the longer a trainee has been umemployed, the more likely he is to smoke ($p < 0.10$).

In the second column of Table 5 are the credit score results. Here we find a strong relationship between one's discount factor and one's credit score but no relationship with being present-biased.[4] Here a standard deviation increase in δ is associated with a 0.121 standard deviation increase in one's credit score ($p < 0.01$). Along with the fact that δ is a strong predictor of credit scores and β is not, we also see that the two other measures of time preference do not significantly correlate with credit scores. Cognitive skills, age, and ethnicity do however, appear to be important: a standard deviation increase in numeracy increases credit scores by 0.080 standard deviations ($p < 0.10$), a similar increase in age raises one's score by nearly a quarter of a standard deviation ($p < 0.01$) and African Americans in our sample tend to have scores that are approximately 50 points lower than the baseline ($p < 0.01$).

The results for body mass index are reported in column (3). Contrary to Chabris et al., (2008), we find no link between time preferences and BMI. We do find a few demographic correlates: being more numerate is associated with higher BMI ($p < 0.05$), non-native English speakers have lower BMI ($p < 0.05$), being unemployed longer is associated with higher BMI ($p < 0.05$) and having military experience is a strong predictor of lower BMI ($p < 0.01$).

Overall, our results on coterminous outcomes suggest that there are a number of significant and interesting bivariate correlations, although many disappear when all our time preference measures are placed in the same regression. When forced to compete, β does well to predict a situation in which weakness of will is surely at play (smoking) but it is δ that seems to perform best.

4.2   Subsequent Outcomes

The administrative records also provide data on subsequent training and job performance. This data is summarized at the bottom of Table 1. For example, we know whether trainees completed their training or "washed out".[5] We also know who completed training but left

---

smoking - and like Bickel et al., (1999) and Mitchell (1999) we find a link to being present-biased.

[4] Our results are not contrary to Meier and Sprenger (2010) who find a strong link between being present-biased and having more credit card debt because more debt does not necessarily translate into lower credit scores unless one does not pay the debt on time.

[5] "Washing Out" here includes both quitting before finishing training, and also failing to pass one of the required phases of training.

the job later, during the two year follow-up data collection period. As it happens, this part of for-hire trucking, long-haul "truckload," has had high turnover rates since the deregulation of 1980. The American Trucking Associations surveys member companies, and until the recent recession had never recorded an average turnover rate among large carriers like the cooperating firm of less than 100% (Burks, et al., 2008). Why such high turnover? This segment of the trucking industry is essentially perfectly competitive, so firms don't have rents to share. The pay is on piece rates (by the mile), and it is hard to earn much at first, given the modest starting rate. The job involves long (on the order of 60) and irregular hours of work each week, and drivers are away from home for two weeks or more at a time, plus drivers have to continually adjust their time use to account for hours of service regulations, shipper and consignee demands, traffic conditions, and weather (Burks, et al., 2010).

In this context turnover might just be a matter of drivers optimally searching for a good match. But the training contract changes the interpretation. When the driver has posted a significant bond that becomes due and payable in full upon exit for any reason before twelve months, even modestly impatient trainees will find it in their interests to continue on the job for the remainder of the year. Therefore, we restrict our analysis to tenure lengths of up to 12 months.

Returning to the bottom of Table 1, despite the strong incentives, two-thirds of the trainees did not complete the 12 months of employment after training that would have canceled their training debt. Considering typical reasons for leaving, eleven percent left before even finishing the driver training program and in many cases drivers simply walked away from the job, abandoning their rigs. In our sample eventually going absent without leave (AWOL) was the outcome for ten percent of the trainees.

To examine the raw relationships between time preferences and subsequent outcomes, in Figure 2 we again break our four time preference measures at the median but now look at the effect of being relatively impatient (or not) on ensuing training and job performance. Note that because the probability of separation for any reason is so high, we graph *not* leaving instead so that all the bars have a common scale. In the first graph we see that the impatience survey does not significantly predict any of the subsequent job outcomes. Although all the β correlations appear to go in the "correct" directions: present biased trainees are less likely to stay on the job, washout more often, and are more likely to go AWOL, only the general separation result is significant ($p = 0.07$) in this simple, first pass, analysis. The results look similar, if not a bit stronger, in the third, δ graph. Here patient trainees are more likely to stay on the job ($p = 0.10$) and are less likely to go AWOL ($p = 0.01$). In this simple bivariate context, the red button measure of impulsivity also performs poorly. In the fourth graph, only the fact that red button clickers seem somewhat more likely to washout of training is

close to significant ($p = 0.11$).

In Table 6 we force the four time preference measures to compete "head-to-head". In column (1) which reports probit marginal effects we see that only present bias predicts leaving for any reason ($p < 0.10$). Reducing present bias by a standard deviation corresponds to a subsequent 0.065 standard deviation decrease in the likelihood of job separation. Cognitive skills also appear to continue to matter here: a standard deviation increase in IQ is associated with a 0.169 standard deviation decrease in the likelihood of leaving ($p < 0.01$). Lastly, African Americans tend to leave significantly more often ($p < 0.05$).

The effects of time preferences on leaving because one did not finish training are presented in column (2) of Table 6. Here we see that only the number of red button clicks *does not* predict washing out of the training program. A standard deviation increase in survey-measured impatience increases the chances of washing out by 0.043 standard deviations ($p < 0.05$). We also see that both β and δ predict, and the effects are as one would expect: a standard deviation reduction in present bias predicts a 0.081 standard deviation reduction in washing out of training ($p < 0.05$) and the effect of δ is similar. Our estimate suggests increasing patience by a standard deviation translates into a 0.066 standard deviation decrease in the likelihood of washing out ($p < 0.05$). Considering the demographics, we find that trainees with higher IQ are less likely to wash out of training ($p < 0.01$) and older, minorities and those who have been unemployed longer are significantly more likely to washout.

When we consider going AWOL from the job in column (3) we see that only the discount factor predicts and the sign of the effect is consistent with intuition. A standard deviation increase in δ, suggests a 0.100 standard deviation decrease in the probability of going AWOL ($p < 0.01$). Other factors that predict going AWOL include cognitive ability and age. Trainees with higher numeracy scores go AWOL less often ($p < 0.05$) as do older trainees ($p < 0.05$).

To summarize, looking at subsequent job performance where we can be more confident of the causal nature of our estimates we see even starker results. Here neither the survey nor our impulsivity experiment predict outcomes with any regularity. Only β and δ seem to predict across domains. Reconsidering the effects presented in Figure 2, it seems that adding δ to our estimates of the effects of β helps - the effect of present bias appears to be clearer only after controlling for one's baseline level of patience.

4.3   Robustness Checks

We also tested whether two alternative formulations of our analysis would significantly change our conclusions. First, to give our new delay of gratification instrument another opportunity

to perform better we created an indicator for clicking the red button any number of times and substituted it into the specifications in Tables 5 and 6. The outcomes of this exercise appear in Table 7. Overall, we find, essentially, the same results - the red button experiment does not compete well when β and δ are included. The only noteworthy difference is that the indicator predicts smoking slightly better than the number of clicks ($p < 0.01$).

Given the historical dominance of the exponential discounting model in economics, we also used all our choice data to reestimate δ under the assumption of no present bias (i.e., restricting $\beta = 1$). In Table 8 we examine how well simple exponential discounting does compared to the results of Tables 5 and 6 that use the quasi-hyperbolic model. As one can see, the exponential model does well, predicting three of the six outcomes, however, this is one less than was previously predicted by δ. Without controlling for present-bias, the exponential discount factor is partially confounded and unable to predict washing out of training. In addition, restricting attention to the coterminous outcomes, one sees that not controlling for β leads the effects of δ to be overstated. While the difference is not too large for credit scores or BMI, in the first column it is clear that the simple discount factor absorbs a considerable amount of variation that should be attributed to present bias.

5   Conclusion

While there is a considerable literature in both economics and psychology that attempts to measure time preferences and link them to important outcomes, there is no strong sense in this literature as to which measures perform best. In most cases, preferences are measured only using a single instrument and results tend only to be published when that measure predicts. With a large sample we test the ability of four different measures (picked to be representative of current methods) to predict outcomes both when they are asked to do so one at a time and when they are forced to compete head-to-head. Considering first bivariate relationships, we find that all four measures predict to some degree; however, while the simple discount factor (δ) predicts four of our six outcomes and present bias (β) predicts three, our new, adult, version of Mischel's test of impulsivity is associated with only two coterminous outcomes and our impatience survey comes close just once. These results overlap to a large degree with the results that we get when we include all the measures in the same regression model and control for a number of other factors. In the end, it appears that the now familiar (β, δ) model of preferences first offered by Phelps and Pollak (1968) is the most robust predictor of outcomes. Further, our regression results suggest that it might also be the case that gathering just δ or β is a mistake, we find that the prediction of outcomes is more robust when both measures are included.

14

6 References

Ameriks, J., Caplin, A., Leahy, J. and Tyler, T., (2007). Measuring self-control problems. American Economic Review, 97(3), 966-972.

Ashraf, N., Karlan, D. and Yin, W., (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. Quarterly Journal of Economics, 121(2), 635-672.

Barsky, R., Juster, F. T., Kimball, M. and Shapiro, M., (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. Quarterly Journal of Economics, 112, 537-579.

Bickel, W., Odum, A. and Madden, G., (1999). Impulsivity and cigarette smoking: Delay discounting in current, never and ex-smokers. Psychopharmacology, 146, 447-454.

Burks, S., Belzer, M., Kwan, Q., Pratt, S., Shackelford, S., (2010). Trucking 101: An industry primer. Transportation Research Circular Number E-C146. Washington, DC, Transportation Research Board.

Burks, S., Carpenter, J., Goette, L., Monaco, K., Porter, K. et al., (2008). Using behavioral economic field experiments at a firm. In S. Bender, J. Lane, K. Shaw, F. Andersson and T. von Wachter (Eds.), The analysis of firms and employees. (pp. 45-106). Chicago: The University of Chicago Press.

Burks, S., Carpenter, J., Goette, L. and Rustichini, A., (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. Proceedings of the National Academy of Sciences, 106(19), 7745-7750.

Chabris, C., Laibson, D., Morris, C., Schuldt, J. and Taubinsky, D., (2008). Individual laboratory-measured discount rates predict field behavior. Journal of Risk and Uncertainty, 37(2), 237-269.

Coller, M. and Williams, M. B., (1999). Eliciting individual discount rates. Experimental Economics, 2(2), 107-127.

Dickman, S., (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. Journal of Personality and Social Psychology, 58(1), 95-102.

Eckel, C., Johnson, C. and Montmarquette, C., (2005). Saving decisions of the working poor: Short- and long-term horizons. In J. Carpenter, G. Harrison and J. List (Eds.), Field experiments in economics. (pp. 219-260). Amsterdam: JAI/Elsevier.

Eckel, C., Johnson, C., Montmarquette, C. and Rojas, C., (2007). Debt aversion and the demand for loans for postsecondary education. Public Finance Review, 35(2), 233-262.

Economic and Statistics Group (2007). Truckload line-haul driver turnover quarterly annualized rates. Trucking Activity Report, 15(3).

Fisher, I., (1930). The theory of interest. New York: Macmillan.

Frederick, S., Loewenstein, G. and O'Donoghue, T., (2002). Time discounting and time preference: A critical review. Journal of Economic Literature, 40(2), 351-401.

Harrison, G. W., Lau, M. I. and Williams, M. B., (2002). Estimating individual discount rates in denmark: A field experiment. American Economic Review, 92(5), 1606-1617.

Jenkins, C. D., Rosenman, R. H. and Friedman, M., (1967). Development of an objective psychological test for the determination of the coronary-prone behavior pattern in employed men. Journal of Chronic Diseases, 20(6), 371-379.

Karlan, D. and Zinman, J., (2008). Lying about borrowing. Journal of the European Economic Association, 6(2-3), 510-521.

Keys, B., Mukherjee, T., Seru, A., Vig, V., (2010). Did Securitizaion lead to lax screening? Evidence from subprime loans. Quarterly Journal of Economics, 125(1), 307-362.

Laibson, D., (1997). Golden eggs and hyperbolic discounting. Quarterly Journal of Economics, 62(2), 443-478.

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

Mischel, W., Shoda, Y. and Rodriguez, M. I., (1989). Delay of gratification in children. Science, 244(4907), 933-938.

Mitchell, S., (1999). Measures of impulsivity in cigarette smokers and non-smokers. Psychopharmacology, 146, 455-464.

Phelps, E. S. and Pollak, R. A., (1968). On second-best national saving and game-equilibrium growth. The Review of Economic Studies, 35(2), 185-199.

Picone, G., Sloan, F. and Taylor, D., (2004). Effects of risk and time preference and expected longevity on demand for medical tests. Journal of Risk and Uncertainty, 28(1), 39-53.

Raven, J., Raven, J. C. and Court, J. H., (2003). Manual for raven's progressive matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment.

Samuelson, P. A., (1937). A note on the measurement of utility. Review of Economic Studies, 4, 155-161.

Shoda, Y., Mischel, W. and Peake, P. K., (1990). Predicting adolescent cognitive and social competence from preschool delay of gratification: Identifying diagnostic conditions. Developmental Psychology, 26(6), 978-986.

Zhang, L. and Rashad, I., (2008). Obesity and time preferences: The health consequences of discounting the future. Journal of Biosocial Science, 40(1), 97-113.
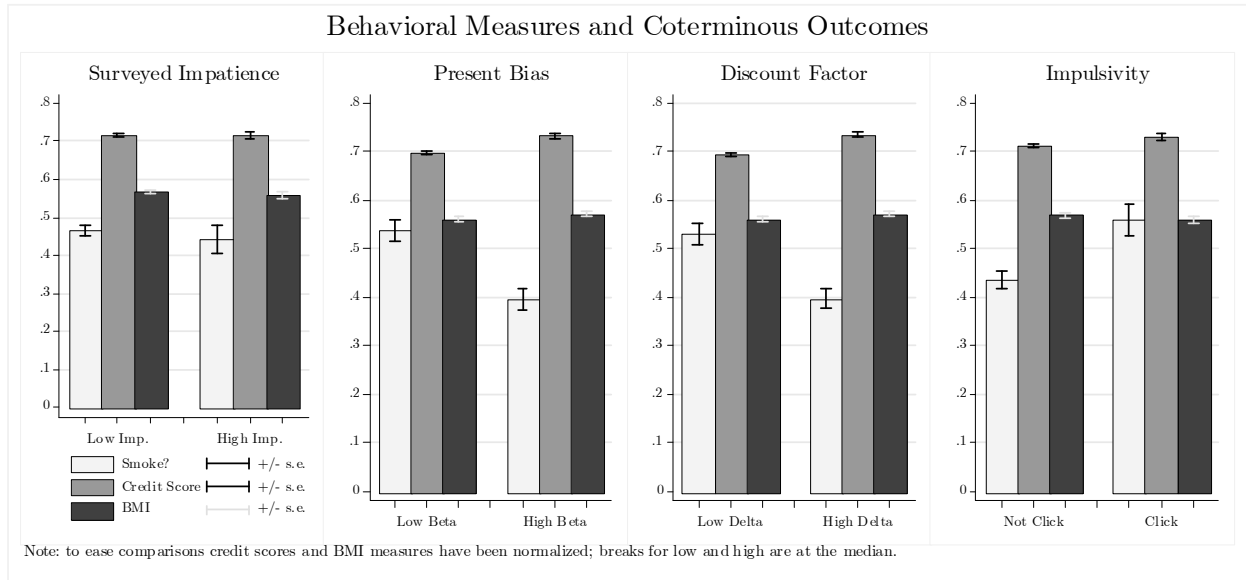
7 Figures and Tables

Figure 1: The Effect of Time Preference on Coterminous Outcomes.
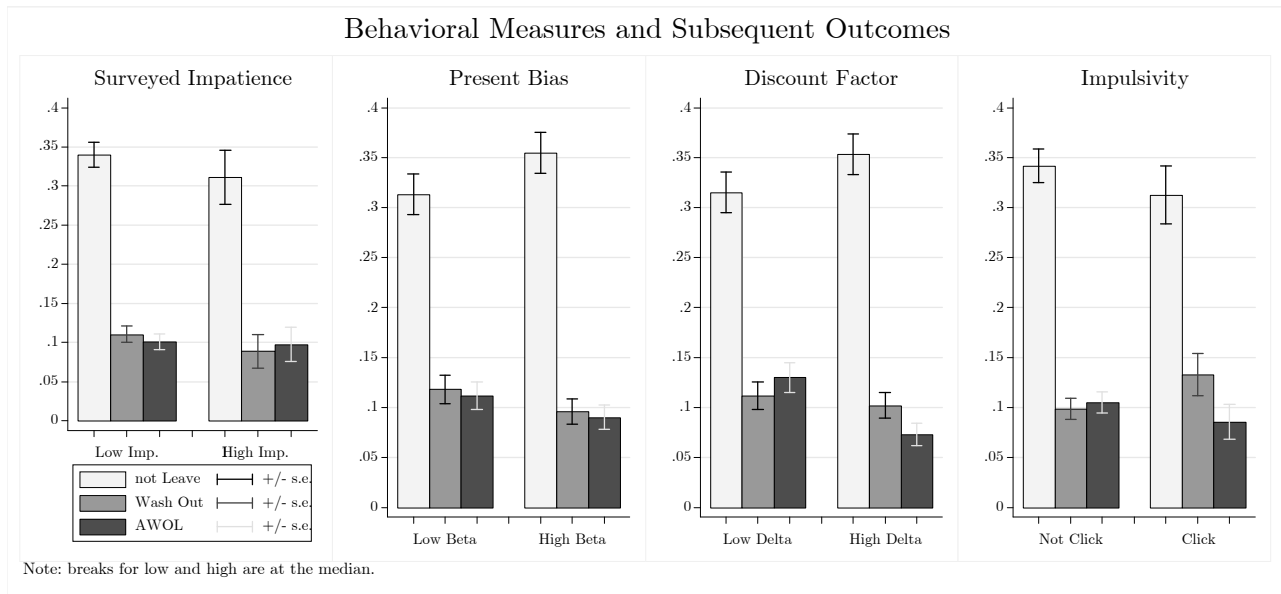
Figure 2: The Effect of Time Preference on Subsequent Outcomes.

Table 1: Descriptive Statistics (N=1069)

Characteristic	Obs.	Mean	Std. Dev.	Min	Max
IQ Score	1013	45.47	7.79	7	60
Numeracy Score	1069	8.42	2.62	0	12
Age (years)	1069	37.43	10.89	21	69
Schooling above High School (I)	1069	0.57	0.49	0	1
English is 2nd Language (I)	1069	0.03	0.18	0	1
African American (I)	1067	0.14	0.35	0	1
Married (I)	1069	0.48	0.50	0	1
Months Unemployed	1068	2.23	4.23	0	24
Family Income less than 30k (I)	1069	0.70	0.45	0	1
Monthly Expenditures above 1.5k (I)	1069	0.45	0.50	0	1
Military Experience (I)	1069	0.28	0.45	0	1
Surveyed Impatience	1069	0.21	0.60	0	6
Present Bias (Beta)	1015	0.90	0.13	0.56	1.07
Discount Factor (Delta)	1015	0.99	0.01	0.97	1.00
Click? (I)	1069	0.24	0.43	0	1
Red Button Clicks	1068	0.38	0.80	0	3
Time Waited	1069	525.09	151.67	7.63	600
Smoke (I)	811	0.46	0.50	0	1
Credit Score	944	588.16	93.21	407	821
Body Mass Index	841	29.71	6.46	17.15	52.46
Leave for any Reason (I)	1069	0.67	0.47	0	1
Washout of Training (I)	1069	0.11	0.31	0	1
AWOL from the Job (I)	955	0.10	0.30	0	1

Table 2: The Distribution of Discounting Types

	$\beta = 1$	$\beta < 1$
$\delta = 1$	52%; $\bar{\delta} = 0.989$, $\bar{\beta} = 0.921$	9%; $\bar{\delta} = 0.980$, $\bar{\beta} = 0.627$
$\delta < 1$	33%; $\bar{\delta} = 0.985$, $\bar{\beta} = 0.967$	6%; $\bar{\delta} = 0.986$, $\bar{\beta} = 0.826$

Table 3: The Determinants of Time Preference

	(1) Survey	(2) Beta	(3) Delta	(4) Click?	(5) Number of Clicks
IQ Score^s	0.029	0.036	0.083**	-0.025	-0.014
	(0.038)	(0.042)	(0.039)	(0.040)	(0.091)
Numeracy Score^s	-0.089	0.185***	0.105***	-0.096***	-0.379***
	(0.073)	(0.043)	(0.041)	(0.039)	(0.098)
Age (years)^s	-0.026	0.091***	0.163***	0.171***	0.453***
	(0.033)	(0.033)	(0.036)	(0.035)	(0.095)
Schooling above High School (I)	0.091*	0.013	0.001**	-0.210**	-0.404***
	(0.049)	(0.009)	(0.001)	(0.095)	(0.137)
English is Second Language (I)	-0.037	0.015	0.004**	0.191	0.454
	(0.081)	(0.023)	(0.002)	(0.243)	(0.308)
African American (I)	-0.028	-0.043***	-0.003***	0.074	0.047
	(0.068)	(0.013)	(0.001)	(0.128)	(0.176)
Married (I)	-0.025	-0.017**	-0.001	0.060	0.089
	(0.032)	(0.009)	(0.001)	(0.094)	(0.141)
Months Unemployed^s	-0.024	0.012	-0.008	0.002	-0.032
	(0.022)	(0.031)	(0.030)	(0.030)	(0.081)
Family Income below 30k (I)	-0.027	-0.008	-0.000	-0.016	-0.055
	(0.053)	(0.009)	(0.001)	(0.101)	(0.149)
Monthly Expenditures above 1.5k (I)	-0.019	-0.001	0.001**	0.104	0.085
	(0.044)	(0.009)	(0.001)	(0.097)	(0.144)
Military Experience (I)	-0.092***	0.007	-0.001	-0.138	-0.047
	(0.029)	(0.010)	(0.001)	(0.107)	(0.162)
Observations	1012	958	958	1012	1012

Marginal effects from OLS, probit and negative binomial estimates; (robust standard errors).
* $p<0.10$, ** $p<0.05$, *** $p<0.01$. ^{s}indicates standardized regression coefficient.

Table 4: Examining the Construct Validity of Clicking

Red Button Clicks	(1)	(2)	(3)	(4)
Present Bias (Beta)	-1.638***			-1.467***
	(0.468)			(0.545)
Discount Factor (Delta)		-17.612**		-6.480
		(7.452)		(8.704)
Impatience (F)			0.077*	0.079
			(0.048)	(0.056)
Constant	0.494	16.411**	-0.959***	6.732
	(0.415)	(7.344)	(0.064)	(8.340)
lnalpha				
Constant	0.796***	0.831***	0.812***	0.787***
	(0.129)	(0.126)	(0.122)	(0.129)
Observations	1014	1014	1068	1014

Marginal effects after negative binomial estimates; (robust standard errors).

* p<0.10, ** p<0.05, *** p<0.01.

Table 5: Time Preferences and Coterminus Outcomes

	(1) Smoke?	(2) Credit Score	(3) BMI
Impatience (F)[s]	-0.014	0.038	-0.014
	(0.044)	(0.031)	(0.031)
Present Bias (Beta)[s]	-0.100**	0.016	0.024
	(0.044)	(0.038)	(0.039)
Discount Factor (Delta)[s]	-0.091**	0.121***	-0.002
	(0.042)	(0.037)	(0.042)
Red Button Clicks[s]	0.033	0.051	-0.023
	(0.042)	(0.033)	(0.033)
IQ Score[s]	0.008	0.011	-0.003
	(0.049)	(0.042)	(0.044)
Numeracy Score[s]	-0.057	0.080*	0.089**
	(0.050)	(0.043)	(0.045)
Age (years)[s]	-0.067	0.235***	-0.022
	(0.043)	(0.039)	(0.039)
Schooling above High School (I)	-0.060	5.339	-0.166
	(0.041)	(6.708)	(0.493)
English is Second Language (I)	-0.065	9.250	-2.463***
	(0.102)	(15.609)	(1.004)
African American (I)	-0.243***	-49.646***	0.526
	(0.053)	(8.268)	(0.667)
Married (I)	-0.066*	1.849	0.633
	(0.039)	(6.229)	(0.478)
Months Unemployed[s]	0.060*	-0.015	0.097**
	(0.037)	(0.037)	(0.040)
Family Income below 30k (I)	0.049	-11.200	-0.206
	(0.042)	(6.954)	(0.512)
Monthly Expenditures above 1.5k (I)	-0.076*	5.303	0.477
	(0.040)	(6.564)	(0.487)
Military Experience (I)	0.055	-1.466	-1.756***
	(0.044)	(6.885)	(0.466)
Observations	754	845	782

Marginal effects after probit estimates and 2SLS; (robust standard errors).

* p<0.10, ** p<0.05, *** p<0.01. [s]indicates standardized regression coefficient.

Table 6: Time Preferences and Subsequent Outcomes

	(1) Leave?	(2) Wash Out?	(3) AWOL?
Impatience (F)^s	0.032	0.047**	-0.007
	(0.042)	(0.022)	(0.036)
Present Bias (Beta)^s	-0.065*	-0.081**	0.043
	(0.039)	(0.034)	(0.035)
Discount Factor (Delta)^s	0.030	-0.066**	-0.100***
	(0.038)	(0.034)	(0.036)
Red Button Clicks^s	-0.032	-0.016	-0.035
	(0.035)	(0.027)	(0.037)
IQ Score^s	-0.169***	-0.094***	0.014
	(0.046)	(0.033)	(0.038)
Numeracy Score^s	-0.050	-0.030	-0.094**
	(0.043)	(0.036)	(0.042)
Age (years)^s	-0.045	0.129***	-0.064*
	(0.038)	(0.032)	(0.039)
Schooling above High School (I)	-0.001	0.017	-0.005
	(0.034)	(0.019)	(0.022)
English is Second Language (I)	-0.089	0.036	0.007
	(0.091)	(0.052)	(0.062)
African American (I)	0.092**	0.135***	0.040
	(0.044)	(0.037)	(0.035)
Married (I)	-0.040	-0.025	0.025
	(0.032)	(0.018)	(0.021)
Months Unemployed^s	0.052	0.051**	0.028
	(0.033)	(0.025)	(0.031)
Family Income below 30k (I)	-0.038	-0.019	-0.004
	(0.035)	(0.022)	(0.023)
Monthly Expenditures above 1.5k (I)	-0.004	0.009	-0.003
	(0.034)	(0.019)	(0.021)
Military Experience (I)	-0.007	-0.014	-0.025
	(0.036)	(0.020)	(0.023)
Observations	958	958	853

Marginal effects after probit estimates; (robust standard errors).

* p<0.10, ** p<0.05, *** p<0.01. ^{s}indicates standardized regression coefficient.

Table 7: Clicking any Number of Times and Outcomes

	(1) Smoke?	(2) Credit Score	(3) BMI	(4) Leave?	(5) Wash Out?	(6) AWOL?
Impatience (F)^s	-0.014	0.038	-0.015	0.031	0.047**	-0.006
	(0.044)	(0.031)	(0.031)	(0.042)	(0.022)	(0.036)
Present Bias (Beta)^s	-0.099**	0.014	0.025	-0.063*	-0.080**	0.045
	(0.044)	(0.038)	(0.039)	(0.039)	(0.034)	(0.035)
Discount Factor (Delta)^s	-0.088**	0.124***	-0.001	0.031	-0.066*	-0.102***
	(0.043)	(0.037)	(0.042)	(0.038)	(0.034)	(0.035)
Click? (I)	0.100***	0.055	-0.009	-0.005	-0.019	-0.035
	(0.041)	(0.034)	(0.035)	(0.035)	(0.026)	(0.031)
Controls Included?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	754	845	782	958	958	853

Marginal effects after probit estimates and OLS; (robust standard errors).
* p<0.10, ** p<0.05, *** p<0.01. ^{s}indicates standardized regression coefficient.

Table 8: Exponential Discounting and Outcomes

	(1) Smoke?	(2) Credit Score	(3) BMI	(4) Leave?	(5) Wash Out?	(6) AWOL?
Impatience (F)^s	-0.021	0.038	-0.012	0.027	0.041**	-0.001
	(0.044)	(0.031)	(0.030)	(0.042)	(0.021)	(0.036)
Exponential Discount Factor (Delta)^s	-0.133***	0.121***	0.010	-0.002	0.021	-0.073**
	(0.036)	(0.031)	(0.036)	(0.032)	(0.028)	(0.030)
Red Button Clicks^s	0.037	0.051	-0.024	-0.029	-0.011	-0.040
	(0.042)	(0.033)	(0.033)	(0.035)	(0.027)	(0.037)
Controls Included?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	754	845	782	958	958	853

Marginal effects after probit estimates and OLS; (robust standard errors).
* p<0.10, ** p<0.05, *** p<0.01. ^{s}indicates standardized regression coefficeint.